

THE MÖBIUS FUNCTION IS STRONGLY ORTHOGONAL TO NILSEQUENCES

BEN GREEN AND TERENCE TAO

ABSTRACT. We show that the Möbius function $\mu(n)$ is strongly asymptotically orthogonal to any polynomial nilsequence $(F(g(n)\Gamma))_{n \in \mathbb{N}}$. Here, G is a simply-connected nilpotent Lie group with a discrete and cocompact subgroup Γ (so G/Γ is a *nilmanifold*), $g : \mathbb{Z} \rightarrow G$ is a polynomial sequence and $F : G/\Gamma \rightarrow \mathbb{R}$ is a Lipschitz function. More precisely, we show that $|\frac{1}{N} \sum_{n=1}^N \mu(n) F(g(n)\Gamma)| \ll_{F,G,\Gamma,A} \log^{-A} N$ for all $A > 0$. In particular, this implies the *Möbius and Nilsequence conjecture* $MN(s)$ from our earlier paper [8] for every positive integer s . This is one of two major ingredients in our programme in [8] to establish a large number of cases of the *generalised Hardy-Littlewood conjecture*, which predicts how often a collection $\psi_1, \dots, \psi_t : \mathbb{Z}^d \rightarrow \mathbb{Z}$ of linear forms all take prime values. The proof is a relatively quick application of the results in our recent companion paper [9].

We give some applications of our main theorem. We show, for example, that the Möbius function is uncorrelated with any bracket polynomial such as $n\sqrt{3}\lfloor n\sqrt{2} \rfloor$. We also obtain a result about the distribution of nilsequences $(a^n x \Gamma)_{n \in \mathbb{N}}$ as n ranges only over the primes.

1. INTRODUCTION

IMPORTANT REMARK. This paper is intimately tied to, and is intended to be read in conjunction with, the longer companion paper [9], which proves results about the distribution of finite polynomial orbits on nilmanifolds. In particular, we shall make heavy use of the notation and lemmas from that paper.

The aim of this paper is to establish what the authors have been referring to as the *Möbius and Nilsequence conjecture* $MN(s)$, first stated as [8, Conjecture 8.5]. Roughly speaking, this states that the *Möbius function* $\mu(n)$, defined as $(-1)^k$ when n is the product of k distinct primes, and 0 otherwise, is asymptotically strongly orthogonal to any Lipschitz s -step nilsequence $(F(a^n x))_{n \in \mathbb{Z}}$, in the sense that the inner product

$$\mathbb{E}_{n \in [N]} \mu(n) F(a^n x)$$

of these two functions on $[N] := \{1, \dots, N\}$ decays to zero faster than any fixed power of $1/\log N$. Here and in the sequel we use the averaging notation $\mathbb{E}_{x \in X} f(x) := \frac{1}{|X|} \sum_{x \in X} f(x)$ for any finite set X . Recall also that an *Lipschitz s -step nilsequence* is any sequence of the form $F(a^n x)$, where a is an element of an s -step connected and simply connected nilpotent Lie group G , x is an element of the *nilmanifold* G/Γ for some discrete cocompact subgroup $\Gamma \leq G$ of G , and $F : G/\Gamma \rightarrow \mathbb{R}$ is a Lipschitz function.

The difficulty of this conjecture increases with s . The case $s = 0$ of this conjecture is the estimate

$$\mathbb{E}_{n \in [N]} \mu(n) \ll_A \log^{-A} N.$$

The stronger estimate

$$\mathbb{E}_{n \in [N]} \mu(n) \ll e^{-c\sqrt{\log N}}$$

is essentially equivalent to the prime number theorem (with classical error term).

The case $s = 1$ may be reduced by Fourier analysis to the estimate

$$|\mathbb{E}_{n \in [N]} \mu(n) e(\alpha n)| \ll_A \log^{-A} N \quad (1.1)$$

where $e(x) := e^{2\pi i x}$, required to hold uniformly for all $\alpha \in \mathbb{R}$. This was established by Davenport [3] in the 1930s by modifying Vinogradov's method of bilinear forms (or "Type I and Type II sums").

In the case $s = 2$ the conjecture was established by the authors in [7]. For a more complete discussion of the conjecture and the reasons for being interested in it (and in particular, its applications to the generalised Hardy-Littlewood conjecture on the number of solutions to systems of linear equations in which the unknowns are all prime) the reader may refer to the introduction of [7], the first several sections of [8], or any of the expository articles [4, 5, 14, 15].

In this paper we settle the Möbius and Nilsequence conjecture. In fact, we shall prove the marginally stronger result that the Möbius function is asymptotically strongly orthogonal to any *polynomial* nilsequence $(F(g(n)\Gamma))_{n \in \mathbb{Z}}$.

Theorem 1.1 (Main Theorem). *Let G/Γ be a nilmanifold of some dimension $m \geq 1$, let G_\bullet be a filtration¹ of G of some degree $d \geq 1$, and let $g \in \text{poly}(\mathbb{Z}, G_\bullet)$ be a polynomial sequence². Suppose that G/Γ has a Q -rational Mal'cev basis³ \mathcal{X} for some $Q \geq 2$, defining a metric $d_{\mathcal{X}}$ on G/Γ . Suppose that $F : G/\Gamma \rightarrow [-1, 1]$ is a Lipschitz function. Then we have the bound*

$$|\mathbb{E}_{n \in [N]} \mu(n) F(g(n)\Gamma)| \ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \log^{-A} N$$

for any $A > 0$ and $N \geq 2$. The implied constant is ineffective.

Remarks. By specialising to the linear case $g(n) := a^n h$ for some $a, h \in G$ (and using the existence of Q -rational Mal'cev bases, see [9, Proposition A.9]), Theorem 1.1 immediately implies the *Möbius and nilsequences conjecture* [8, Conjecture 8.5]. In fact it gives a somewhat more precise result, since the dependence on Q and $\|F\|_{\text{Lip}}$ is given quite explicitly. For the application of Theorem 1.1 in [8], however, knowledge of these dependencies is not necessary.

The ineffectivity of the bound in Theorem 1.1 already occurs for sufficiently large A in the 1-step case (which, as mentioned before, is essentially (1.1)), and is ultimately due to the well-known ineffective bounds on Siegel zeroes. On the other hand, the remainder of the argument is effective, and so any effective bound for Siegel's theorem would imply effective bounds for Theorem 1.1. In particular, this would be the case if one assumed GRH. In fact, in that case it is not difficult to see from modifying the arguments below

¹In other words, $G_\bullet = (G_i)_{i=0}^d$ where $G = G_0 \subset G_1 \subset \dots \subset G_d$ is a descending sequence of Lie groups and $[G_i, G_j] \subset G_{i+j}$ for all $i, j \geq 0$, with the convention that G_i is trivial for $i > d$; see [9, Definition 1.2].

²A sequence $g : \mathbb{Z} \rightarrow G$ lies in $\text{poly}(\mathbb{Z}, G_\bullet)$ if $\partial_{h_1} \dots \partial_{h_i} g$ takes values in G_i for all $h_1, \dots, h_i \in \mathbb{Z}$ and $i \geq 0$, where $\partial_h g(n) := g(n+h)g(n)^{-1}$; see [9, Definition 1.11] and the ensuing discussion.

³The notion of a Q -rational Mal'cev basis is defined in [9, Definition 2.6] and the construction of the metric $d_{\mathcal{X}}$ is given in the same section.

that we can replace the logarithmic decay $\log^{-A} N$ by polynomial decay N^{-c} for some $c > 0$ depending only on d and m .

The authors learnt in [9] that it is in many ways more natural to consider the class of polynomial sequences $\text{poly}(\mathbb{Z}, G_\bullet)$ rather than simply the class of linear sequences $n \mapsto a^n x$. This is ultimately due to the stability of the polynomial class under a wide variety of operations, such as pointwise multiplication. On the other hand, these two categories are certainly closely related (and are, in some sense, equivalent): see [12] for further discussion.

ACKNOWLEDGEMENTS. The first author is partly supported by a Leverhulme Prize. The second author is supported by a grant from the Macarthur Foundation and by NSF grant DMS-0649473.

2. REDUCING TO THE EQUIDISTRIBUTED CASE

To prove Theorem 1.1, we will apply [9, Theorem 1.19] to decompose g as a product $\varepsilon g' \gamma$ where ε is “smooth”, γ is “rational” and g' is highly equidistributed in some closed subgroup $G' \subseteq G$. We will recall the precise statement shortly.

In this section we shall show how the rather harmless factors ε and γ in the above factorisation may be eliminated, and then make an additional reduction to the case $\int_{G/\Gamma} F = 0$ (using the Haar measure on G/Γ , of course). This leaves us with the task of proving an “equidistributed” case of Theorem 1.1: see Proposition 2.1 below.

For the rest of the paper, all constants c, C , including those in the asymptotic notation \ll and $O()$, are allowed to depend on m and d . Different occurrences of the letters c, C may represent different constants; typically we will have $0 < c \ll 1 \ll C < \infty$. For ease of notation we drop the subscript whenever Lipschitz norms are mentioned, so $\|F\|_{\text{Lip}}$ becomes simply $\|F\|$.

Recall from [9, Definition 1.3(v)] that a sequence $(g(n)\Gamma)_{n \in [N]}$ in a nilmanifold is *totally δ -equidistributed* if we have

$$|\mathbb{E}_{n \in P} F(g(n)\Gamma)| \leq \delta \|F\| \quad (2.1)$$

for all Lipschitz functions $F : G/\Gamma \rightarrow \mathbb{C}$ with $\int_{G/\Gamma} F = 0$ and all arithmetic progressions $P \subseteq [N]$ of length at least δN .

In the next section we shall establish the following result about the lack of correlation of Möbius with equidistributed nilsequences.

Proposition 2.1 (Möbius is orthogonal to equidistributed sequences). *Let $m \geq 0$, $d \geq 1$ be integers and let $N \geq 1$ be an integer parameter which is sufficiently large depending on m and d . Let δ , $0 < \delta < 1/2$, and $Q \geq 2$ be real parameters. Let G/Γ be an m -dimensional nilmanifold, and suppose that G_\bullet is a filtration of degree d . Suppose that G/Γ has a Q -rational Mal'cev basis \mathcal{X} adapted to the filtration G_\bullet . Let $g \in \text{poly}(\mathbb{Z}, G_\bullet)$ and suppose that $(g(n)\Gamma)_{n \in [N]}$ is totally δ -equidistributed. Then for any function $F : G/\Gamma \rightarrow \mathbb{R}$ with $\int_{G/\Gamma} F = 0$ and for any arithmetic progression $P \subseteq [N]$ of size at least N/Q , we have the bound*

$$|\mathbb{E}_{n \in [N]} \mu(n) 1_P(n) F(g(n)\Gamma)| \ll \delta^c Q \|F\| \log N.$$

The proof of Proposition 2.1 proceeds via the method of Type I/II sums, which is also known as the method of bilinear forms. This is the same method that one might use to tackle the “minor arcs” case of (1.1), where α is not close to a rational with small denominator. We will describe it in detail in the next section. Our task for the remainder of this section is to reduce Theorem 1.1 to Proposition 2.1.

Proof that Proposition 2.1 implies Theorem 1.1. We start with a brief overview. The main ingredient of this argument is [9, Theorem 1.19], that is to say the factorization $g = \varepsilon g' \gamma$ mentioned above. In addition to that we require estimates for sums of the type $\mathbb{E}_{n \in [N]} \mu(n) 1_P(n)$, where $P \subseteq [N]$ is a progression. After standard harmonic analysis, such bounds ultimately depend on results about the zeros of L -functions $L(s, \chi)$, and as such this is analysis of the same type as would be used to establish the “major arc” cases of (1.1). Finally, a fair amount of what might be called “quantitative nil-linear algebra” is required to keep track of the various nilmanifolds and Lipschitz functions involved in the argument. Here we draw repeatedly on the material assembled in [9, Appendix A] for this purpose; we encourage the reader to gloss over these essentially routine issues on a first reading.

We now turn to the details. We allow all implied constants to depend on m and d .

Let the hypotheses be as in Theorem 1.1. To simplify the notation slightly we will also assume that $\|F\| \geq 1$; the case $\|F\| < 1$ can easily be deduced from that case. By dividing out by $\|F\|$ we may in fact normalize and assume that $\|F\| = 1$.

We may of course take $A \geq 1$. We may also assume that $Q \leq \log N$, since the claim is vacuously true otherwise; thus \mathcal{X} is now a log N -rational Mal’cev basis. By increasing A if necessary, it will suffice to show an estimate of the form

$$|\mathbb{E}_{n \in [N]} \mu(n) F(g(n) \Gamma)| \ll_A \log^{-A+O(1)} N. \quad (2.2)$$

Let B be a parameter (depending on A) to be specified later. We may assume that N is sufficiently large depending on A, B . By [9, Theorem 1.19] (with $M_0 := \log N$) we can find an integer M ,

$$\log N \leq M \ll \log^{O_B(1)} N,$$

a rational subgroup $G' \subseteq G$, a Mal’cev basis \mathcal{X}' for G'/Γ' (where $\Gamma' := G \cap \Gamma$) in which each element is an M -rational combination (see [9, Definition 1.21]) of the elements of \mathcal{X} , and a decomposition

$$g = \varepsilon g' \gamma \quad (2.3)$$

into polynomial sequences $\varepsilon, g', \gamma \in \text{poly}(\mathbb{Z}, G_\bullet)$ with the following properties:

- (i) $\varepsilon : \mathbb{Z} \rightarrow G_\bullet$ is (M, N) -smooth (see [9, Definition 1.22] for a definition);
- (ii) $g' : \mathbb{Z} \rightarrow G'$ takes values in G' , and the finite sequence $(g'(n) \Gamma')_{n \in [N]}$ is totally M^{-B} -equidistributed in G'/Γ' , using the metric $d_{\mathcal{X}'}$ on G'/Γ' ;
- (iii) $\gamma : \mathbb{Z} \rightarrow G$ is M -rational (see [9, Definition 1.21]), and $(\gamma(n) \Gamma)_{n \in \mathbb{Z}}$ is periodic with period $1 \leq q \leq M$.

From (2.3) we have

$$\mathbb{E}_{n \in [N]} \mu(n) F(g(n) \Gamma) = \mathbb{E}_{n \in [N]} \mu(n) F(\varepsilon(n) g'(n) \gamma(n) \Gamma). \quad (2.4)$$

The sequence $(\gamma(n) \Gamma)_{n \in \mathbb{Z}}$ is periodic with some period q , $1 \leq q \leq M$. For each $j = 0, 1, \dots, q-1$ let $\gamma_j := \{\gamma(j)\}$ be the fractional part of $\gamma(j)$ with respect to Γ , thus

$\gamma_j \Gamma = \gamma(j) \Gamma$ and all the coordinates $\psi_{\mathcal{X}}(\gamma_j)$ lie in $[0, 1)$. This construction is described in [9, Lemma A.14].

Now by [9, Lemma A.12], the coordinates $\psi_{\mathcal{X}}(\gamma(j))$ lie in $\frac{1}{M'} \mathbb{Z}^m$ for some $M' \ll M^{O(1)}$. Since $\gamma_j = \gamma(j) \eta$ for some η with integer coordinates, it follows from [9, Lemma A.3] that the coordinates $\psi_{\mathcal{X}}(\gamma_j)$ are rationals with height $\ll M^{O(1)}$.

We now take advantage of the periodicity of $\gamma(n) \Gamma$ to split the right-hand side of (2.4) as

$$\sum_{j=0}^{q-1} \mathbb{E}_{n \in [N]} \mu(n) 1_{n \equiv j \pmod{q}} F(\varepsilon(n) g'(n) \gamma_j \Gamma); \quad (2.5)$$

By the right-invariance of d , the (M, N) -smoothness of ε (see [9, Definition 1.21]) and the 1-Lipschitz bound on F we see that

$$\begin{aligned} |F(\varepsilon(n) g'(n) \gamma_j \Gamma) - F(\varepsilon(n_0) g'(n) \gamma_j \Gamma)| &\leq d_{\mathcal{X}}(\varepsilon(n) g'(n) \gamma_j, \varepsilon(n_0) g'(n) \gamma_j) \\ &= d_{\mathcal{X}}(\varepsilon(n_0), \varepsilon(n)) \\ &\leq \log^{-A} N. \end{aligned}$$

whenever if $|n - n_0| \leq \frac{N}{M \log^A N}$. Hence if we split each progression $n \equiv j \pmod{q}$ into further progressions $P_{j,k}$ for $k = O(M \log^A N)$, each having diameter at most $\frac{N}{M \log^A N}$, we see that (2.5) is equal to

$$\sum_{j,k} \mathbb{E}_{n \in [N]} \mu(n) 1_{P_{j,k}}(n) F(a_{j,k} g'(n) \gamma_j \Gamma) + O(\log^{-A} N). \quad (2.6)$$

Here each $a_{j,k} := \varepsilon(n_{0,j,k})$ for some $n_{0,j,k} \in P_{j,k}$; by the definition of what it means for $\varepsilon : \mathbb{Z} \rightarrow G$ to be (M, N) -smooth (i.e. [9, Definition 1.21]), it follows that $d_{\mathcal{X}}(a_{j,k}, \text{id}_G) \leq M$ and hence, by [9, Lemma A.4], that

$$|\psi_{\mathcal{X}}(a_{j,k})| \ll M^{O(1)}. \quad (2.7)$$

If N is sufficiently large depending on A and B then $N \geq 10M \log^A N$ (say), and this partition of $[N]$ may be arranged in such a way that

$$|P_{j,k}| \geq \frac{N}{2qM \log^A N} \geq \frac{N}{2M^2 \log^A N}.$$

Since the number of j is at most M , and the number of k is at most $M \log^A N$, we thus see that to show (2.2) it suffices by the triangle inequality to show that

$$|\mathbb{E}_{n \in [N]} \mu(n) 1_{P_{j,k}}(n) F(a_{j,k} g'(n) \gamma_j \Gamma)| \ll_A M^{-2} \log^{-2A+O(1)} N \quad (2.8)$$

for each j, k .

Fix j, k . Write $H_j := \gamma_j^{-1} G' \gamma_j$ and let $g_j : \mathbb{Z} \rightarrow H_j$ be the sequence defined by $g_j(n) := \gamma_j^{-1} g'(n) \gamma_j$. It is clear that each g_j is a polynomial sequence with coefficients in the filtration $(H_j)_{\bullet} := \gamma_j^{-1} G'_{\bullet} \gamma_j$.

Set $\Lambda_j := \Gamma \cap H_j$ and define functions

$$F_{j,k} : H_j / \Lambda_j \rightarrow [-1, 1]$$

by the formula

$$F_{j,k}(x \Lambda_j) := F(a_{j,k} \gamma_j x \Gamma).$$

Then (2.8) can be rewritten as

$$|\mathbb{E}_{n \in [N]} \mu(n) 1_{P_{j,k}}(n) F_{j,k}(g_j(n) \Lambda_j)| \ll_A M^{-2} \log^{-2A+O(1)} N. \quad (2.9)$$

Suppose for the moment that $F_{j,k}$ were a constant function. Recall that $P_{j,k}$ has common difference $q \leq M$. We may thus apply Proposition A.2 (with A replaced by a sufficiently large exponent A' depending on A and B) to obtain the desired claim, since $M \ll \log^{O_B(1)} N$. Therefore we may subtract off the mean of $F_{j,k}$ and assume without loss of generality that $\int_{H_j/\Lambda_j} F_{j,k} = 0$. This may cause $F_{j,k}$ to take values in $[-2, 2]$ rather than $[-1, 1]$, but we can easily counter this trivial issue by dividing $F_{j,k}$ by two.

In a moment we shall use Proposition 2.1 to estimate the terms appearing here. Before doing that we record quantitative rationality properties of the nilmanifold H_j/Λ_j , as well as a Lipschitz bound on $\|F_{j,k}\|$.

Claim. There is a Mal'cev basis \mathcal{Y}_j for H_j/Λ_j adapted to the filtration $(H_j)_\bullet$ such that each \mathcal{Y}_j is an M^C -rational combination of the X_i . With respect to the metric $d_{\mathcal{Y}_j}$ on H_j/Λ_j induced by this basis, the polynomial sequence $g_j \in \text{poly}(\mathbb{Z}, (H_j)_\bullet)$ is $M^{-cB+O(1)}$ -totally equidistributed for some $c > 0$ depending only on m, d , and we have $\|F_{j,k}\| \leq M^{O(1)}$.

Proof. We shall apply suitable combinations of the lemmas in [9, Appendix A]. The existence of \mathcal{Y}_j follows from Proposition A.9 and Lemma A.13 of [9] together with the fact that each γ_j has rational coordinates with height $M^{O(1)}$. Now the map $x \mapsto F(a_{j,k} \gamma_j x \Gamma)$ on G/Γ has Lipschitz constant at most $M^{O(1)}$ by [9, Lemma A.5] and the bounds $|\psi_{\mathcal{X}}(a_{j,k})|, |\psi_{\mathcal{X}}(\gamma_j)| \leq M^{O(1)}$. The final statement of the claim, and the statement about the quantitative equidistribution of g_j , now follow from [9, Lemma A.17]. \square

Let us now apply Proposition 2.1 to (2.9). We apply the proposition with parameters (which we distinguish using tildes) as follows: $\tilde{G} := H_j$, $\tilde{\Gamma} := \Lambda_j$, $\tilde{G}_\bullet := (H_j)_\bullet$, $\tilde{g} := g_j$, $\tilde{X} := \mathcal{Y}_j$, $\tilde{Q} := M^{O(1)}$, $\tilde{F} := F_{j,k}$ and $\tilde{\delta} := M^{-cB+O(1)}$. We quickly see that (2.9) is bounded by $O\left(M^{-cB+O(1)} \log^{O(A)} N\right)$. Choosing B sufficiently large depending on A , we obtain (2.9) as claimed. \square

3. THE EQUIDISTRIBUTED CASE: TYPE I AND II SUMS

In this section we establish Proposition 2.1 using Vinogradov's method of Type I and II sums in the form due to Vaughan [16]. More precisely, we will use the following proposition.

Proposition 3.1 (Method of Type I/II sums). *Let $f : \mathbb{N} \rightarrow \mathbb{C}$ be a function with $\|f\|_\infty \leq 1$ such that*

$$|\mathbb{E}_{N < n \leq 2N} \mu(n) \overline{f(n)}| \geq \varepsilon$$

for some $\varepsilon > 0$. Then one of the following statements holds:

- (Type I sum is large) *There exists an integer $1 \leq K \leq N^{2/3}$ such that*

$$|\mathbb{E}_{N/k < w \leq 2N/k} f(kw)| \gg (\varepsilon / \log N)^{O(1)} \quad (3.1)$$

for $\gg (\varepsilon / \log N)^{O(1)} K$ integers k such that $K < k \leq 2K$.

- (Type II sum is large) *There exist integers K, W with $\frac{1}{2}N^{1/3} \leq K \leq 4N^{2/3}$ and $N/4 \leq KW \leq 4N$, such that*

$$|\mathbb{E}_{K < k, k' \leq 2K} \mathbb{E}_{W \leq w, w' < 2W} f(kw) \overline{f(k'w)} \overline{f(kw')} f(k'w')| \gg (\varepsilon / \log N)^{O(1)}. \quad (3.2)$$

Proof. This is [7, Proposition 4.2], specialised to the case $U = V = N^{1/3}$, and with certain explicit exponents replaced by unspecified constants $O(1)$. \square

We now begin the proof of Proposition 2.1. As before we may normalise so that $\|F\| = 1$. From this and the mean zero assumption, we see in particular that

$$|F(x)| \leq \text{diam}(G/\Gamma) \ll Q^{O(1)} \quad (3.3)$$

for all $x \in G/\Gamma$ (the diameter bound here is [9, Lemma A.16]).

If $\delta \leq 1/N$ then by (2.1) we have $|F(g(n)\Gamma)| \leq \delta$ for all $n \in [N]$, and the claim is trivial, so we may assume that $\delta > 1/N$. By increasing δ if necessary (and shrinking c) we thus see that we may assume that

$$\delta > N^{-\sigma} \quad (3.4)$$

for any fixed small constant $\sigma > 0$ depending only on m, d .

The basic idea, which will become clearer upon reading the details, is to make good use of the fact that one may test the quantitative equidistribution properties of a polynomial nilsequence on G/Γ by passing to the abelianisation $(G/\Gamma)_{\text{ab}}$, a phenomenon referred to in [9, Theorem 2.9] as the “quantitative Leibman Dichotomy” (cf. [12]). The abelian issues that one must then deal with are of a very similar nature to those involved in dealing with exponential sums such as $\mathbb{E}_{n \in [N]} \mu(n) e(p(n))$, where $p : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$ is an ordinary polynomial. Rather than quote results from the existing literature on this problem it is easier for us to invoke various lemmas from [9], which were stated and proved in a language which is helpful for the present paper.

Let $\varepsilon := \delta^{c_1} Q \log N$, for a constant c_1 to be specified later. We may assume that $\varepsilon < 1$, otherwise the claim is trivial from (3.3) and the triangle inequality. In particular, we have

$$Q, \log N \leq \delta^{-c_1}$$

and we will use these estimates frequently in the sequel to absorb any polynomial factors in Q or $\log N$ into a power of δ^{-c_1} .

Suppose for contradiction that Proposition 2.1 failed for these parameters. We then apply Proposition 3.1 with $f(n) := 1_P(n) F(g(n)\Gamma)$ and ε as above, concluding that either (3.1) or (3.2) holds. We deal with these two cases in turn.

The Type I case. Suppose that (3.1) holds. Thus there are $\gg \delta^{O(c_1)} K$ values of $k \in (K, 2K]$ such that

$$|\mathbb{E}_{N/k < w \leq 2N/k} 1_P(kw) F(g(kw)\Gamma)| \gg \delta^{O(c_1)}.$$

Let l denote the common difference of P ; since $|P| \geq N/Q$, we must have $1 \leq l \leq Q$. Splitting into progressions with common difference l , we see that for some $b \pmod{l}$ and

for $\gg \delta^{O(c_1)} K$ values of $k \in (K, 2K]$ we have

$$\left| \sum_{\substack{N/k < w \leq 2N/k \\ w \equiv b \pmod{l}}} 1_P(kw) F(g(kw)\Gamma) \right| \gg \delta^{O(c_1)} \frac{N}{kl}.$$

Setting $w = b + lw'$, this may be rewritten as

$$\left| \sum_{w' \in I_k} F(g(k(b + lw')\Gamma) \right| \gg \delta^{O(c_1)} \frac{N}{kl}, \quad (3.5)$$

where $I_k \subseteq [\frac{N}{2kl} - 1, \frac{N}{kl}]$ is an interval.

For each value of k for which this holds, consider the sequence $g_k : \mathbb{Z} \rightarrow G$ defined by $g_k(n) := g(kn)$ and also the sequence $\tilde{g}_k : \mathbb{Z} \rightarrow G$ defined by $\tilde{g}_k(n) = g(k(b + ln))$. It follows from [9, Corollary 6.8] that $g_k, \tilde{g}_k \in \text{poly}(\mathbb{Z}, G_\bullet)$. Now (3.5) implies that $(\tilde{g}_k(n)\Gamma)_{n \in [N_k]}$ fails to be $\delta^{O(c_1)}$ -equidistributed in G/Γ , where $N_k \sim N/kl$.

It follows from [9, Theorem 2.9] that there is a nontrivial horizontal character $\psi_k : G \rightarrow \mathbb{R}/\mathbb{Z}$ (i.e. a continuous homomorphism from G to \mathbb{R}/\mathbb{Z} which annihilates Γ) with magnitude $|\psi_k| \ll \delta^{-O(c_1)}$ such that

$$\|\psi_k \circ \tilde{g}_k\|_{C^\infty[N_k]} \ll \delta^{-O(c_1)}.$$

Recall from [9, Definition 2.10] that the $C^\infty[N]$ -norm of a polynomial $p : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ expanded in binomial coefficients as

$$p(n) = \alpha_0 + \alpha_1 \binom{n}{1} + \cdots + \alpha_d \binom{n}{d}, \quad (3.6)$$

is defined by

$$\|p\|_{C^\infty[N]} := \sup_{1 \leq j \leq d} N^j \|\alpha_j\|_{\mathbb{R}/\mathbb{Z}}.$$

By [9, Lemma 8.4] (specialised to the single-parameter case $t = 1$), there is some $q_k \ll \delta^{-O(c_1)}$ such that

$$\|q_k \psi_k \circ g_k\|_{C^\infty[N_k]} \ll \delta^{-O(c_1)}.$$

Pigeonholing in the possible choices of $q_k \psi_k$, we may find some ψ with $0 < |\psi| \ll \delta^{-O(c_1)}$ such that

$$\|\psi \circ g_k\|_{C^\infty[N_k]} \ll \delta^{-O(c_1)} \quad (3.7)$$

for $\gg \delta^{O(c_1)} K$ values of $k \in (K, 2K]$.

Write

$$\psi \circ g(n) = \beta_d n^d + \cdots + \beta_0. \quad (3.8)$$

Then

$$\psi \circ g_k(n) = \beta_d k^d n^d + \cdots + \beta_0. \quad (3.9)$$

We would like to use this and (3.7) to conclude that the coefficients $k^j \beta_j$ are close to being integer (or rational with small denominator). This will follow from a simple lemma.

Lemma 3.2. *Suppose that $p : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a polynomial of the form $p(n) = \beta_d n^d + \cdots + \beta_0$. Then there is some $q \geq 1$, $q = O(1)$, such that $\|q\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll N^{-j} \|p\|_{C^\infty[N]}$ for $j = 1, \dots, d$.*

Proof. Consider the representation (3.6) which is used to define the $C^\infty[N]$ -norm. Observing that β_j can be written as a linear combination of $\alpha_j, \dots, \alpha_d$ with rational coefficients of height $O(1)$, the result follows upon clearing denominators. \square

From (3.7), (3.9) and Lemma 3.2 we see that there is some $q \geq 1$, $q = O(1)$, such that

$$\|qk^j\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}(N/K)^{-j} \quad (3.10)$$

for $j = 1, 2, \dots, d$ and for at least $\delta^{O(c_1)}K$ values of $k \in (K, 2K]$.

Fix j , $1 \leq j \leq d$. To pass from the j^{th} powers k^j to more general integers we shall need the following Waring-type result.

Lemma 3.3. *Let $K \geq 1$ be an integer, and suppose that $S \subseteq [K]$ is a set of size αK . Suppose that $t \geq 2^j + 1$. Then $\gg_{j,t} \alpha^{2t} K^j$ integers in the interval $[tK^j]$ can be written in the form $k_1^j + \dots + k_t^j$, $k_1, \dots, k_t \in S$.*

Proof. It is a well-known consequence of Hardy and Littlewood's asymptotic formula for Waring's problem (see e.g. [17]) that the number of solutions to

$$x_1^j + \dots + x_t^j = M, \quad x_1, \dots, x_t \in [K]$$

is $\ll_{j,t} K^{t-j}$ uniformly in M provided that $t \geq 2^j + 1$. (In fact, by subsequent work, such a result is known for much smaller values of t when j is large.) Let $X = \{k^j : k \in S\}$ and let $r(n)$ be the number of representations of n as the sum of t elements of X . Then by the Cauchy-Schwarz inequality and the preceding remarks we have

$$\alpha^{2t} K^{2t} = \left(\sum_n r(n) \right)^2 \leq |tX| \sum_n r(n)^2 \ll_j |tX| K^{2t-j},$$

which implies the result. \square

By (3.10) and Lemma 3.3 it follows that

$$\|ql\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}(K/N)^j$$

for $\gg \delta^{O(c_1)}K^j$ values of $l \in [10^d K^j]$.

The following lemma, which is [9, Lemma 3.2], may be applied to this situation.

Lemma 3.4 (Strongly recurrent linear functions are highly non-diophantine). *Let $\alpha \in \mathbb{R}$, $0 < \sigma < 1/2$, and $0 < \mu \leq \sigma/2$, and let $I \subseteq \mathbb{R}/\mathbb{Z}$ be an interval of length μ such that $\alpha n \in I$ for at least σN values of $n \in [N]$. Then there is some $k \in \mathbb{Z}$ with $0 < |k| \ll \sigma^{-O(1)}$ such that $\|k\alpha\|_{\mathbb{R}/\mathbb{Z}} \ll \mu \sigma^{-O(1)}/N$.* \square

Let us attempt to apply this lemma with $\sigma \gg \delta^{O(c_1)}$ and $\mu \ll \delta^{-O(c_1)}(K/N)^j$. If N is sufficiently large and the exponent σ in (3.4) is sufficiently small, we see using the bound $K/N \leq N^{-1/3}$ that the hypotheses of the lemma are satisfied and that such an application is permissible. The conclusion is that there is some q' , $1 \leq q' \ll \delta^{-O(c_1)}$, such that

$$\|qq'\beta_i\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}N^{-i}. \quad (3.11)$$

Writing $\tilde{\psi} := qq'\psi$, it follows from (3.8) and (3.11) that for any n we have the bound

$$\|\tilde{\psi} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}n/N.$$

If $N' := \delta^{C_1} N$ for some sufficiently large C , and if $n \in [N']$, this implies that

$$\|\tilde{\psi} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \leq 1/10. \quad (3.12)$$

Now set $\tilde{F} : G/\Gamma \rightarrow [-1, 1]$ to be the function $\tilde{F} := \eta \circ \tilde{\psi}$, where $\eta : \mathbb{R}/\mathbb{Z} \rightarrow [-1, 1]$ is a function of Lipschitz norm $O(1)$ and mean zero which equals 1 on $[-1/10, 1/10]$. Then we have $\int_{G/\Gamma} \tilde{F} = 0$ and $\|\tilde{F}\| \ll \delta^{-O(c_1)}$. From (3.12), we have

$$|\mathbb{E}_{n \in [N']} \tilde{F}(g(n)\Gamma)| \geq 1 > \delta \|\tilde{F}\|,$$

provided that c_1 is chosen sufficiently small. This is contrary to the assumption that $(g(n)\Gamma)_{n \in [N]}$ is δ -totally equidistributed.

The Type II case. This is in many ways very closely similar to the Type I case, as the reader will see. Recall the situation that (3.2) puts us in (with our choice of ε): there are K, W with $\frac{1}{2}N^{1/3} \leq K \leq 4N^{2/3}$ and $N/4 \leq KW \leq 4N$ such that

$$|\mathbb{E}_{K < k, k' \leq 2K} \mathbb{E}_{W < w, w' \leq 2W} f(kw)f(kw')f(k'w)f(k'w')| \gg \delta^{O(c_1)},$$

where $f(n) = 1_P(n)F(g(n)\Gamma)$. Writing the left-hand side here as

$$\mathbb{E}_{K < k, k' \leq 2K} |\mathbb{E}_{W < w \leq 2W} f(kw)f(k'w)|^2,$$

we see that there are $\gg \delta^{O(c_1)} K^2$ pairs $(k, k') \in (K, 2K]^2$ such that

$$|\mathbb{E}_{W < w \leq 2W} f(kw)f(k'w)| \gg \delta^{O(c_1)}.$$

Written out in full, for each such pair (k, k') we have

$$|\mathbb{E}_{W < w \leq 2W} 1_P(kw)1_P(k'w)F(g(kw)\Gamma)F(g(k'w)\Gamma)| \gg (\varepsilon/\log N)^{O(1)}.$$

Writing l for the common difference of P (thus $1 \leq l \leq Q$) we see that there is some $b \pmod l$ such that for $\gg (\varepsilon/\log N)^{O(1)} K^2$ pairs (k, k') we have

$$\sum_{\substack{W < w \leq 2W \\ w \equiv b \pmod l}} 1_P(kw)1_P(k'w)F(g(kw)\Gamma)F(g(k'w)\Gamma) \gg \delta^{O(c_1)} \frac{W}{l}.$$

Setting $w = lw' + b$, this may be written as

$$\left| \sum_{w' \in I_{k,k'}} F(g(k(b+lw'))\Gamma)F(g(k'(b+lw'))\Gamma) \right| \gg \delta^{O(c_1)} \frac{W}{l}, \quad (3.13)$$

where $I_{k,k'} \subseteq (\frac{W}{l} - 1, \frac{2W}{l}]$ is an interval. Since $1 \leq l \leq Q$, which is bounded by a small power of N , and $W \gg N^{1/3}$, this is contained in $[\frac{W}{2l}, \frac{2W}{l}]$.

For each k, k' for which this holds, consider the sequence $g_{k,k'} : \mathbb{Z} \rightarrow G \times G$ defined by $g_{k,k'}(n) = (g(kn), g(k'n))$, and also the sequence $\tilde{g}_{k,k'} : \mathbb{Z} \rightarrow G \times G$ defined by $\tilde{g}_{k,k'}(n) = (g(k(b+ln)), g(k'(b+ln)))$. It follows from [9, Corollary 6.8] that $g_{k,k'}, \tilde{g}_{k,k'} \in \text{poly}(\mathbb{Z}, G_\bullet \times G_\bullet)$. Now from (3.13) we see that the sequence $(\tilde{g}_{k,k'}(n)(\Gamma \times \Gamma))_{n \in [N_{k,k'}]}$ fails to be $\delta^{O(c_1)}$ -equidistributed in $(G/\Gamma) \times (G/\Gamma)$, for some $N_{k,k'} \in [\frac{W}{2l}, \frac{2W}{l}]$.

It follows from [9, Theorem 2.9] that there is a nontrivial horizontal character $\psi_{k,k'} : G \times G \rightarrow \mathbb{R}/\mathbb{Z}$ with $|\psi_k| \ll \delta^{-O(c_1)}$ such that

$$\|\psi_{k,k'} \circ \tilde{g}_{k,k'}\|_{C^\infty[N_{k,k'}]} \ll \delta^{-O(c_1)}.$$

By [9, Lemma 8.4] there is some $q_{k,k'} \ll \delta^{-O(c_1)}$ such that

$$\|q_{k,k'} \psi_{k,k'} \circ g_{k,k'}\|_{C^\infty[N_{k,k'}]} \ll \delta^{-O(c_1)}.$$

Pigeonholing in the possible choices of $q_{k,k'} \psi_{k,k'}$, we may find some ψ with $0 < |\psi| \ll \delta^{-O(c_1)}$ such that

$$\|\psi \circ g_{k,k'}\|_{C^\infty[N_{k,k'}]} \ll \delta^{-O(c_1)} \quad (3.14)$$

for $\gg \delta^{O(c_1)} K^2$ pairs $k, k' \in (K, 2K]$.

Write $\psi = \psi_1 \oplus \psi_2$, where $\psi_1, \psi_2 : G \rightarrow \mathbb{R}/\mathbb{Z}$ are horizontal characters, not both zero. If

$$\psi_1 \circ g(n) = \beta_d n^d + \cdots + \beta_0$$

and

$$\psi_2 \circ g(n) = \beta'_d n^d + \cdots + \beta'_0$$

then

$$\psi \circ g_{k,k'}(n) = (\beta_d k^d + \beta'_d k'^d) n^d + \cdots + (\beta_0 + \beta'_0),$$

By Lemma 3.2 and (3.14) there is some $1 \leq q \ll \delta^{-O(c_1)}$ such that

$$\|q(k^j \beta_j + k'^j \beta'_j)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)} N_{k,k'}^{-j} \ll \delta^{-O(c_1)} (K/N)^j$$

for $j = 1, 2, \dots, d$ and for $\gg \delta^{O(c_1)} K^2$ pairs $k, k' \in (K, 2K]$.

Suppose, without loss of generality, that $\psi_1 \neq 0$. Selecting some k' that occurs in $\gg \delta^{O(c_1)} K$ of the pairs k, k' and subtracting, we see that

$$\|q k^j \beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)} (K/N)^j \quad (3.15)$$

for $\gg \delta^{O(c_1)} K$ values of $k \in (-K, K)$. Using the bounds $K \gg N^{1/3}$ and (3.4) it follows that we may ignore the contribution of $k = 0$, that is to say (3.15) holds for $\gg \delta^{O(c_1)} K$ values of $k \in [1, K]$.

Remark. Note carefully that (3.15) carries no information when $k = 0$. In our treatment of Type I sums there was no need for a lower bound on K , but such an assumption is essential if one has any desire to bound Type II sums.

The estimate (3.15) is identical to (3.10). We may now repeat the arguments used to obtain a contradiction to (3.10) in Type I case. The proof of Proposition 2.1 and thus Theorem 1.1 is now complete. \square

The main business of the paper is now complete. In the next section we give a brief discussion of how our argument compares with the classical Hardy-Littlewood method. After that we give a number of applications of Theorem 1.1.

4. REMARKS ON A NILPOTENT HARDY-LITTLEWOOD METHOD

It may be of interest to interpret our method in terms of the “major and minor arcs” terminology of the Hardy-Littlewood method. Recall that to prove Davenport’s estimate

$$|\mathbb{E}_{n \in [N]} \mu(n) e(\alpha n)| \ll_A \log^{-A} N$$

one divides into two cases: the *major arcs* where α is close to a rational with small denominator, and the *minor arcs* where it is not. The major arcs are handled using

L -function technology as in Appendix A, and the minor arcs are handled using Type I/II sums as in Proposition 3.1.

Suppose that we are considering the sum

$$\mathbb{E}_{n \in [N]} \mu(n) F(g(n)\Gamma),$$

where $\int_{G/\Gamma} F = 0$. Decompose g as a product $\varepsilon g' \gamma$ where ε is smooth, γ is rational and g' is highly equidistributed on some subgroup G' . Then one might think of g as a “major arc” nilsequence if $G' = \{\text{id}_G\}$, and as “minor arc” if G' is nontrivial.

To justify this terminology, observe that one may interpret $e(\alpha n)$ as $F(g(n)\Gamma)$, where $G/\Gamma = \mathbb{R}/\mathbb{Z}$, $g : \mathbb{Z} \rightarrow \mathbb{R}$ is the polynomial sequence $g(n) = \alpha n$ and the Lipschitz function F , taking values in the unit ball of the complex plane, is simply $e(\theta)$.

If $\alpha = \frac{a}{q} + \varepsilon$, where ε is small, then the decomposition $g = \varepsilon g' \gamma$ will be given by $\varepsilon(n) = \varepsilon n$, $g'(n) = \text{id}_G$ and $\gamma(n) = an/q$ and so this does indeed correspond to a “major arc nilsequence”.

If α is not close to a rational with small denominator then $g(n)$ will already be highly equidistributed on \mathbb{R}/\mathbb{Z} , and so the decomposition $g = \varepsilon g' \gamma$ has $\varepsilon = \gamma = \text{id}_G$ and $g' = g$. Thus $G' = \mathbb{R}$ is nontrivial and this corresponds to a “minor arc nilsequence”.

5. ON BRACKET POLYNOMIALS

By a *bracket polynomial* we mean an object formed from the scalar field \mathbb{R} and the indeterminate n using finitely many instances of the standard arithmetic operations $+$, \times together with the integer part operation $\lfloor \cdot \rfloor$ and the fractional part operation $\{ \cdot \}$. The following are all bracket polynomials: $n^2 + n\sqrt{2}$, $n\sqrt{2} \lfloor n\sqrt{3} \rfloor$ and $\{n^3\sqrt{2} + n^7 \lfloor n\sqrt{5} \rfloor + \sqrt{7}\}$. One may associate a notion of *complexity* to any bracket polynomial $p(n)$, this being (for instance) the least number of operations $+$, \times , $\lfloor \cdot \rfloor$, $\{ \cdot \}$ required to write down p . In view of the relation $\{x\} + \lfloor x \rfloor = x$, it is not strictly speaking necessary to retain both the integer and fractional part operations, but we do so here for convenience. Dispensing with one of them would slightly alter the definition of complexity.

The following remarkable theorem of Bergelson and Leibman [2] demonstrates a close link between bracket polynomials and nilmanifolds. If G/Γ is a nilmanifold with Mal'cev basis \mathcal{X} then recall from [9, Lemma A.14] that the coordinate map $\psi : G \rightarrow \mathbb{R}^m$ provides an identification between G/Γ and $[0, 1)^m$. Write τ_1, \dots, τ_m for the individual coordinate maps from G/Γ to $[0, 1)$, that is to say τ_i is the composition of ψ with the map $(t_1, \dots, t_m) \mapsto t_i$.

Theorem 5.1 (Bergelson-Leibman). *The functions of the form $n \mapsto \{p(n)\}$, where p is a bracket polynomial, coincide with the functions of the form $n \mapsto \tau_i(g(n)\Gamma)$, where G/Γ is a nilmanifold equipped with a Mal'cev basis \mathcal{X} and $g : \mathbb{Z} \rightarrow G$ is a polynomial map with coefficients in some filtration G_\bullet . The rationality of \mathcal{X} , the dimension of G , the degree of g and the rationality of G_\bullet may all be bounded in terms of the complexity of p , and conversely the complexity of p may be bounded in terms of these quantities.*

In fact, Bergelson and Leibman prove a number of rather refined variants of this type of result, and they also give a comprehensive and edifying discussion of bracket polynomials in general. At first glance it appears that one might immediately combine

Theorem 5.1 with Theorem 1.1 to obtain a result about the correlation of the Möbius function with bracket polynomials. There is a serious catch, however: the coordinate functions τ_i are not continuous on the nilmanifold G/Γ . Furthermore, as observed by Bergelson and Leibman, there are bracket polynomials which *cannot* be written in the form $F(g(n)\Gamma)$ for a continuous F . Indeed the results of Leibman [12] on the distribution of $(g(n)\Gamma)_{n \in \mathbb{Z}}$ imply that the sequence $(F(g(n)\Gamma))_{n \in \mathbb{Z}}$ cannot have isolated values, yet there are bracket polynomials which do. A simple example is $\lfloor 1 - \{n\sqrt{2}\} \rfloor$, which is zero except when $n = 0$.

One does nonetheless feel that the discontinuities of τ_i are “mild”, as this function is continuous on that part of G/Γ which is identified with $(0, 1)^m$. However, the sequence $(g(n)\Gamma)_{n \in \mathbb{Z}}$ may well concentrate on a highly singular subset of G/Γ , as we discussed at length in [9]. Thus a certain amount of further work is required to obtain the expected result, which is the following.

Theorem 5.2 (Möbius and bracket polynomials). *Suppose that $p(n)$ is a bracket polynomial and that $\Psi : [0, 1] \rightarrow [-1, 1]$ is a Lipschitz function. Then we have the estimate*

$$\mathbb{E}_{n \in [N]} \mu(n) \Psi(\{p(n)\}) \ll_{A, \Psi} \log^{-A} N,$$

where the implied constant depends only on A , Ψ and the complexity of p (but is ineffective).

We shall illustrate how this theorem may be deduced from Theorem 1.1 by discussing two related special cases. We will then sketch the details that are required in order to write down a complete proof. The authors plan to include a complete proof of Theorem 5.2 in a future publication.

Both special cases will take place on the Heisenberg nilmanifold G/Γ , where

$$G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}, \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.$$

Computations with Mal'cev bases in this setting were given in [7, Appendix B] and then again in [9, §5], where we took

$$e_1 = \exp(X_1) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, e_2 = \exp(X_2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, e_3 = \exp(X_3) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We briefly recall some of the computations carried out in somewhat more detail in that paper; in any case the proofs are nothing more than computations with 3×3 matrices. The coordinate function $\psi : G \rightarrow \mathbb{R}^3$ is then given by the formula

$$\psi \left(\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \right) = (x, y, z - xy),$$

and the element written here is equivalent, under right multiplication by an element of Γ , to the element with coordinates

$$(\{x\}, \{y\}, \{z - xy - \lfloor x \rfloor y\}).$$

Note that this lies inside the fundamental domain $[0, 1)^3$. It follows that, for any $\alpha, \beta \in \mathbb{R}$, we have

$$\{n\beta \lfloor n\alpha \rfloor\} = \tau_3(g(n)\Gamma),$$

where $\tau_3 : G/\Gamma \rightarrow [0, 1)$ is the map into the third coordinate and $g : \mathbb{Z} \rightarrow G$ is the polynomial sequence given by

$$g(n) = \begin{pmatrix} 1 & n\alpha & n^2\alpha\beta \\ 0 & 1 & n\beta \\ 0 & 0 & 1 \end{pmatrix}.$$

This is an explicit example of the representation of a bracket polynomial, in this case $\{n\beta \lfloor n\alpha \rfloor\}$, in the form discussed in Bergelson and Leibman's theorem.

We discuss two different cases.

Case 1. $\alpha = \sqrt{2}$, $\beta = \sqrt{3}$. Then the sequence $(g(n)\Gamma)_{n \in [N]}$ is totally N^{-c} -equidistributed on G/Γ , which makes life rather easy. To prove the equidistribution one may use [9, Theorem 2.9] together with the lower bound

$$\min_{\substack{|k_1|, |k_2|, |k_3| \leq K \\ (k_1, k_2, k_3) \neq (0, 0, 0)}} \|k_1\sqrt{2} + k_2\sqrt{3}\|_{\mathbb{R}/\mathbb{Z}} \gg K^{-C},$$

which follows from the fact that, for any k_3 with $|k_3| \leq K$, $k_1\sqrt{2} + k_2\sqrt{3} + k_3$ satisfies a quartic over \mathbb{Z} with coefficients of size $K^{O(1)}$. Although the function τ_3 is not continuous, it is continuous outside of a subset of G/Γ of measure zero, namely outside of $[0, 1)^3 \setminus (0, 1)^3$. This means that it may be approximated by Lipschitz functions. More precisely, for any fixed Lipschitz function $\Psi : [0, 1] \rightarrow [-1, 1]$ and any $\varepsilon > 0$ one may find functions $F_1, F_2 : G/\Gamma \rightarrow \mathbb{C}$ with $\|F_1\|_\infty, \|F_2\|_\infty \leq 1$, $\|F_1\|_{\text{Lip}}, \|F_2\|_{\text{Lip}} \leq \varepsilon^{-O(1)}$, $|\Psi \circ \tau_3 - F_1| \leq F_2$ pointwise and $\int_{G/\Gamma} F_2 \leq \varepsilon$. From Proposition (2.1) we have

$$\mathbb{E}_{n \in [N]} \mu(n) F_1(g(n)\Gamma) \ll N^{-c},$$

and the uniform distribution of $(g(n)\Gamma)_{n \in [N]}$ implies that

$$\mathbb{E}_{n \in [N]} F_2(g(n)\Gamma) \leq \varepsilon + O(\varepsilon^{-O(1)} N^{-c}).$$

Now we have the bounds

$$\begin{aligned} |\mathbb{E}_{n \in [N]} \mu(n) \Psi(n\sqrt{3} \lfloor n\sqrt{2} \rfloor)| &= |\mathbb{E}_{n \in [N]} \mu(n) \Psi \circ \tau_3(g(n)\Gamma)| \\ &\leq |\mathbb{E}_{n \in [N]} \mu(n) F_1(g(n)\Gamma)| + \mathbb{E}_{n \in [N]} F_2(g(n)\Gamma). \end{aligned}$$

Letting $\varepsilon = N^{-c'}$ for some sufficiently small $c' > 0$, we obtain an effective and much stronger version of Theorem 5.2 in this case, namely the bound

$$\mathbb{E}_{n \in [N]} \mu(n) \Psi(\{n\sqrt{3} \lfloor n\sqrt{2} \rfloor\}) \ll N^{-c}.$$

Case 2. $\alpha = \beta = \sqrt{2}$. Now the sequence $(g(n)\Gamma)_{n \in [N]}$ is manifestly *not* uniformly distributed on G/Γ . In fact g takes values in the one-dimensional subgroup $G' \subseteq G$ defined by

$$G' = \left\{ \begin{pmatrix} 1 & x & x^2/2 \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}.$$

The preceding argument breaks down. One could appeal to Theorem 1.1 instead of Proposition 2.1, but the problem comes when one tries to control the term

$$\mathbb{E}_{n \in [N]} F_2(g(n)\Gamma).$$

Without knowing something more about the relation between the support properties of F_2 and the orbit $(g(n)\Gamma)_{n \in [N]}$, it is not possible to control this term.

In the case at hand $(g(n)\Gamma)_{n \in [N]}$ is N^{-c} -equidistributed in the nilmanifold G'/Γ' where $\Gamma' := \Gamma \cap G$. Topologically and algebraically this nilmanifold is nothing more than \mathbb{R}/\mathbb{Z} , but one should note carefully that the Haar measure on this nilmanifold is not the same as the measure induced from the Haar measure on G . This may be used to “explain” the observation that $n\sqrt{2}\lfloor n\sqrt{2} \rfloor$ is not uniformly distributed modulo one; see [2] for further details.

Inside G/Γ , G'/Γ' may be identified with the union of two segments

$$\left\{ \begin{pmatrix} 1 & x & x^{2/2} \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq x < 1 \right\} \cup \left\{ \begin{pmatrix} 1 & x & (1+x^2)/2 \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq x < 1 \right\},$$

and this makes it clear that the induced map $\tau_3 : G'/\Gamma' \rightarrow [0, 1)$ is continuous away from a single point. By an analysis very similar to the preceding one it may once again be shown that

$$\mathbb{E}_{n \in [N]} \mu(n) \Psi(\{n\sqrt{2}\lfloor n\sqrt{2} \rfloor\}) \ll N^{-c}$$

for any fixed Lipschitz function $\Psi : [0, 1] \rightarrow [-1, 1]$.

Amongst examples of the form $n\beta\lfloor n\alpha \rfloor$ there is a third distinct case, typified by $\alpha = \beta = 2^{1/3}$. We leave the analysis of this to the reader.

Sketch proof of the general case of Theorem 5.2. By Theorem 5.1, the result of Bergelson and Leibman, it suffices to show, for any fixed Lipschitz function $\Psi : [0, 1] \rightarrow [-1, 1]$, that

$$\mathbb{E}_{n \in [N]} \mu(n) (\Psi \circ \tau_i)(g(n)\Gamma) \ll_A \log^{-A} N.$$

Here, the notation and parameters are as described in Theorem 5.1. Now τ_i is continuous outside the set $[0, 1)^m \setminus (0, 1)^m$, which has zero measure in G/Γ . The issue lies in understanding how the orbit $(g(n)\Gamma)_{n \in [N]}$ interacts with this.

Now the main results of [9] allow us to get a handle on this situation. Consider in particular the decomposition of g as $\varepsilon g' \gamma$ which was obtained in [9, Theorem 1.19]. Recall that $\varepsilon : \mathbb{Z} \rightarrow G$ is slowly varying, $\gamma : \mathbb{Z} \rightarrow G$ is rational and $g' : \mathbb{Z} \rightarrow G'$ is such that $(g'(n)\Gamma')_{n \in [N]}$ is totally equidistributed. For a full proof of Theorem 5.2 one would naturally need to specify appropriate quantitative parameters here. Suppose for simplicity that $\varepsilon = \gamma = \text{id}_G$ (this was, in fact, the case in the two examples above).

Choose a Mal'cev basis for G'/Γ' with coordinate map $\psi' : G' \rightarrow \mathbb{R}^{m'}$. Then G'/Γ' may be identified with the region $\psi'^{-1}([0, 1)^{m'}) \subseteq G$, and in this way we think of the coordinate function τ_i as a function on G'/Γ' . Write $\tilde{\tau}_i$ for the corresponding function on $[0, 1)^{m'}$. It can be shown, making extensive use of the results of [9, Appendix A], that $\tilde{\tau}_i$ is continuous outside of a *piecewise polynomial set* of positive codimension, that is to say outside of a finite union of sets each of which is defined by some polynomial inequalities $a \leq P(t_1, \dots, t_{m'}) < b$ and at least one nontrivial polynomial equation $Q(t_1, \dots, t_{m'}) = c$. Related matters are discussed at greater length in [2]; in the two examples we discussed, these piecewise polynomial sets were rather simple. These sets are certainly well-behaved enough that τ_i may be approximated using Lipschitz functions F_1 and F_2 as in our treatment of the bracket polynomial $n\sqrt{3}\lfloor n\sqrt{2} \rfloor$, and in this way one may use Theorem 1.1 to obtain the desired bound

$$\mathbb{E}_{n \in [N]} \mu(n) (\Psi \circ \tau_i)(g'(n)\Gamma) \ll_A \log^{-A} N.$$

If $G' \neq \{\text{id}\}$ then one may in fact use Proposition 2.1 to obtain the stronger bound of N^{-c} , as in the examples.

If ε and γ are not trivial it is even more complicated to write down a fully rigorous argument, but conceptually things are not much harder at all. The introduction of the smooth function $\varepsilon(n)$ has a rather benign effect; if n ranges over an interval of length $\delta'N$, for suitably small $\delta' = \delta'(\delta)$, the discontinuities of the functions $x \mapsto \tau_i(\varepsilon(n)x\Gamma)$ are all contained inside a “nice” set of measure at most δ , and one may proceed much as before. All one need do, then, is split the range $[N]$ into suitably short intervals of this type.

The introduction of γ may be handled much as it was in the proof of Theorem 1.1. One splits each of the intervals from the previous paragraph into progressions P_j with the same (small) common difference q such that $\gamma(n)\Gamma$ is constant and equal to $\gamma_j\Gamma$ on P . One then works with the conjugated sequences $\gamma_j^{-1}g'(n)\gamma_j$ as we did at the end of §2. \square

We conclude by remarking on some variants and generalizations of Theorem 5.2. If p_1, \dots, p_M are bracket polynomials and $F : (\mathbb{R}/\mathbb{Z})^M \rightarrow \mathbb{C}$ is a smooth function then one could establish the estimate

$$\mathbb{E}_{n \in [N]} \mu(n) F(\{p_1(n)\}, \dots, \{p_M(n)\}) \ll_A \log^{-A} N$$

by Fourier decomposition of F and Theorem 5.2. One could, if desired, restrict the range of the average to some fixed subprogression $P \subseteq [N]$ by the standard technique of approximating the cutoff $1_P(n)$ by a smoother function $\tilde{1}_P(n)$ and then developing this as a Fourier expansion.

6. THE LIOUVILLE FUNCTION

Everything we have proved for the Möbius function also holds for the Liouville function $\lambda : \mathbb{N} \rightarrow \{-1, 1\}$, defined to be the unique completely multiplicative function such that $\lambda(p) = -1$ for all primes p . This function is related to the Möbius function via the identity

$$\lambda(n) = \sum_{r:r^2|n} \mu(n/r^2).$$

Thus, with the notation and assumptions of Theorem 1.1, we have

$$|\mathbb{E}_{n \in [N]} \lambda(n) F(g(n)\Gamma)| \ll \sum_{1 \leq r \leq \sqrt{N}} \frac{1}{r^2} |\mathbb{E}_{m \in [N/r^2]} \mu(m) F(g(r^2 m)\Gamma)|.$$

Now by [9, Corollary 6.8] $m \mapsto g(r^2 m)$ is a polynomial sequence with coefficients in the same filtration G_\bullet as g , and so we have the bound

$$|\mathbb{E}_{m \in [N/r^2]} \mu(m) F(g(r^2 m)\Gamma)| \ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \log^{-A}(N/r^2)$$

uniformly in r , so long as $N/r^2 \geq 2$. Summing over r we obtain

$$\begin{aligned} |\mathbb{E}_{n \in [N]} \lambda(n) F(g(n)\Gamma)| &\ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \left(\sum_{r \leq \sqrt{N}/2} \frac{1}{r^2} \log^{-A}(N/r^2) \right. \\ &\quad \left. + \sum_{\sqrt{N}/2 < r \leq \sqrt{N}} \frac{1}{r^2} \right) \\ &\ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \log^{-A} N. \end{aligned}$$

This is precisely Theorem 1.1, but with λ taking the place of μ . In a similar fashion, all of the results of the preceding section concerning bracket polynomials may now also be deduced with λ in place of μ .

7. A RECURRENCE RESULT ALONG THE PRIMES

In this section we derive the following result. Here p_1, p_2, p_3, \dots is the sequence of primes.

Theorem 7.1 (Prime return times on a nilmanifold). *Suppose that G/Γ is a nilmanifold and that $g \in G$ is such that left-multiplication by g is ergodic. Then for every $x \in G/\Gamma$ the sequence $(g^{p_n}x\Gamma)_{n=1,2,\dots}$ is equidistributed in G/Γ in the sense that*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} F(g^{p_n}x\Gamma) = \int_{G/\Gamma} F$$

for all continuous functions $F : G/\Gamma \rightarrow [-1, 1]$.

Remarks. We recall (from discussions in the companion paper [9]) Leon Green's criterion for ergodicity of left-multiplication by g ; this map is ergodic if and only if rotation by $\pi(g)$ is ergodic on the horizontal torus $(G/\Gamma)_{\text{ab}}$, that is to say if and only if the entries of $\pi(g)$ together with 1 are linearly independent over \mathbb{Q} . If this is the case then left-multiplication by any power of g is uniquely ergodic, that is to say

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} F(g^{t_n}x\Gamma) = \int_{G/\Gamma} F \quad (7.1)$$

for all $x \in G/\Gamma$ and for $t = 1, 2, 3, \dots$

Proof of Theorem 7.1. Let w be a large number and set $W := \prod_{p \leq w} p$. Fix a nilmanifold G/Γ and a continuous (and hence Lipschitz) function $F : G/\Gamma \rightarrow [-1, 1]$. Then uniformly in the residues b coprime to W we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \left(\frac{\phi(W)}{W} \Lambda'(Wn + b) - 1 \right) F(g^n x\Gamma) = o_{w \rightarrow \infty}(1), \quad (7.2)$$

where the convergence is uniform in $x \in G/\Gamma$ and $g \in G$. This follows very quickly from [8, Proposition 10.2], which was proved under the assumption of the Möbius and Nilsequences conjectures $\text{MN}(s)$ which we have established in this paper. Recall that $\Lambda'(p) = \log p$ and that $\Lambda'(n) = 0$ if n is not a prime, that is to say Λ' is a modified version of the von Mangoldt function with no support on the prime powers p^2, p^3, \dots . We recall that the proof of (7.2) is quite substantial. One splits the von Mangoldt function Λ in a certain way as the sum of two pieces $\Lambda^\sharp + \Lambda^\flat$. The contribution from the second piece is bounded using the $\text{MN}(s)$ conjecture, and this is not particularly difficult. The

contribution from the first piece is bounded using the machinery of Gowers norms, and here one must estimate the dual Gowers norm of the nilsequence $F(g^n x \Gamma)$ as well as the Gowers norm of objects related to Λ^\sharp . This is a substantial amount of work.

Let us return to the proof at hand. Since (7.2) is uniform in g and x , we may replace g by g^W and x by $g^b x$ to get

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} \left(\frac{\phi(W)}{W} \Lambda'(n) - 1 \right) F(g^n x \Gamma) = o_{w \rightarrow \infty}(1)$$

uniformly for all progressions $P_{b,W} = \{Wn + b : n \in [N]\}$, $b = 0, 1, \dots, W-1$. However it follows from (7.1) that, for fixed b and W ,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} F(g^n x \Gamma) = \int_{G/\Gamma} F.$$

Comparing these last two expressions we obtain

$$\frac{\phi(W)}{W} \lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F + o_{w \rightarrow \infty}(1),$$

uniformly for b coprime to W . Now if b is not coprime to W we obviously have

$$\frac{\phi(W)}{W} \lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} \Lambda'(n) F(g^n x \Gamma) = o_{w \rightarrow \infty}(1)$$

since Λ' is supported on the primes and F is bounded by 1.

Summing over b , one may conclude that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [WN]} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F + o_{w \rightarrow \infty}(1).$$

This is easily seen to imply that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F + o_{w \rightarrow \infty}(1).$$

The left-hand side no longer depends on w , so we may let $w \rightarrow \infty$. Doing so, we obtain

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F.$$

An easy argument using the prime number theorem, noting that $\Lambda'(p_n)$ is essentially $\log N$ for almost all primes p_n , $n \leq N$, concludes the proof. \square

Very straightforward approximation arguments allow one to replace the continuous function F by a function with mild discontinuities. In this way one could prove, for example, that the sequence $p_n \sqrt{3} \lfloor p_n \sqrt{2} \rfloor$ is uniformly distributed modulo one. We leave the details, which are essentially all present in the earlier discussion of $n \sqrt{3} \lfloor n \sqrt{2} \rfloor$, to the reader.

APPENDIX A. MÖBIUS AND PERIODIC FUNCTIONS

In this appendix we give the proof of Proposition A.2. The argument is, quite apart from being completely standard, already contained in [7, Chapter 3]. We nonetheless take the opportunity to recall it here, as we wish to emphasise the fact that the main

input to this part of the argument is information on the zeros of L -functions. Our starting point is the following proposition.

Proposition A.1. *For any $A > 0$ we have*

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{\chi(n)} \ll_A q^{1/2} \log^{-A} N \quad (\text{A.1})$$

for all Dirichlet characters χ to modulus q .

Remark. This follows from the nonexistence of zeros of $L(s, \chi)$ close to the line $\Re s = 1$. For the details, see [10, Prop. 5.29]. As noted in [10, p. 124] there are difficulties involved in applying the standard Perron's formula approach to $\mathbb{E}_{n \in [N]} \mu(n) \chi(n)$ directly, and it is rather easier to first obtain bounds on $\mathbb{E}_{n \in [N]} \Lambda(n) \chi(n)$.

Using standard techniques of harmonic analysis we may obtain the following consequence of Proposition A.1.

Proposition A.2 (Möbius is orthogonal to periodic sequences). *Let $f : \mathbb{N} \rightarrow \mathbb{C}$ be a sequence bounded in magnitude by 1 which is periodic of some period $q \geq 1$. Then we have*

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} \ll_A q \log^{-A} N$$

for all $A > 0$, where the implied constant is ineffective.

Proof. We first establish the estimate under the additional assumption that $f(n)$ vanishes whenever $(n, q) \neq 1$. Then f can be viewed as a function on the multiplicative group $(\mathbb{Z}/q\mathbb{Z})^\times$, and thus has a Fourier expansion

$$f(n) = \sum_{\chi} \hat{f}(\chi) \chi(n), \text{ where } \hat{f}(\chi) := \mathbb{E}_{n \in (\mathbb{Z}/q\mathbb{Z})^\times} f(n) \overline{\chi(n)},$$

with χ ranging over all the characters on $(\mathbb{Z}/q\mathbb{Z})^\times$. Applying Proposition A.1 and the triangle inequality, we conclude

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} \ll_A q^{1/2} \log^{-A} N \left(\sum_{\chi} |\hat{f}(\chi)| \right).$$

But from Cauchy-Schwarz and Plancherel we have

$$\sum_{\chi} |\hat{f}(\chi)| \leq \phi(q)^{1/2} \left(\sum_{\chi} |\hat{f}(\chi)|^2 \right)^{1/2} = \phi(q)^{1/2} (\mathbb{E}_{n \in (\mathbb{Z}/q\mathbb{Z})^\times} |f(n)|^2)^{1/2} = O(\phi(q)^{1/2}),$$

where $\phi(q) := |(\mathbb{Z}/q\mathbb{Z})^\times|$ is the Euler totient function. Since $\phi(q) \leq q$, the claim follows.

Now we consider the general case, in which (n, q) is not necessarily equal to 1 on the support of f . Observe that if $\mu(n)$ is non-zero, then n is square-free, and we can split $n = dm$, where $d = (n, q)$ is square-free (so $\mu^2(d) = 1$) and m is coprime to q . Furthermore we have $\mu(n) = \mu(d)\mu(m)$. We thus obtain the decomposition

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} = \frac{1}{N} \sum_{d|q; \mu^2(d)=1} \mu(d) \sum_{1 \leq m \leq N/d} \mu(m) \overline{f(dm)} 1_{(m,q)=1}. \quad (\text{A.2})$$

The sequence $m \mapsto f(dm) 1_{(m,q)=1}$ is periodic of period q/d and vanishes whenever $(m, q/d) \neq 1$, hence by the preceding arguments

$$\sum_{1 \leq m \leq N/d} \mu(m) \overline{f(dm)} 1_{(m,q)=1} \ll_A \frac{Nq}{d^2} \log^{-A} N.$$

Thus from (A.2) we have

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} \ll_A q \log^{-A} N \sum_{d|q} \frac{1}{d^2} \ll q \log^{-A} N,$$

concluding the proof of Proposition A.2. \square

REFERENCES

- [1] V. Bergelson and I. J. Håland, *Sets of recurrence and generalized polynomials*, Convergence in ergodic theory and probability (Columbus, OH, 1993), 91–110, Ohio State Univ. Math. Res. Inst. Publ., **5**, de Gruyter, Berlin, 1996.
- [2] V. Bergelson and A. Leibman, *Distribution of values of bounded generalized polynomials*, Acta Mathematica **198** (2007), 155–230.
- [3] H. Davenport, *On some infinite series involving arithmetical functions. II*, Quart. J. Math. Oxf. **8** (1937), 313–320.
- [4] B. J. Green *Generalising the Hardy-Littlewood method for primes*, International Congress of Mathematicians. Vol. II, 373–399, Eur. Math. Soc., Zurich, 2006.
- [5] ———, *Three topics in additive prime number theory*, submitted to Current Developments in Mathematics 2007, Harvard.
- [6] B. J. Green and T. C. Tao, *An inverse theorem for the Gowers U^3 -norm*, Proc. Edinburgh Math. Soc. **51** (2008), no. 1, 73–153.
- [7] B. J. Green and T. C. Tao, *Quadratic uniformity of the Möbius function*, to appear in Annales de l’Institut Fourier (Grenoble).
- [8] ———, *Linear equations in primes*, to appear in Annals of Math.
- [9] ———, *The quantitative behaviour of polynomial orbits on nilmanifolds*, submitted together with this paper.
- [10] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, **53**. American Mathematical Society, Providence, RI, 2004. xii+615 pp
- [11] A. Leibman, *Polynomial sequences in groups*, Journal of Algebra **201** (1998), 189–206.
- [12] ———, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory and Dynamical Systems **25** (2005), no. 1, 201–213.
- [13] A. Mal’cev, *On a class of homogeneous spaces*, Izvestiya Akad. Nauk SSSR, Ser Mat. **13** (1949), 9–32.
- [14] T. C. Tao, *Obstructions to uniformity, and arithmetic patterns in the primes*, special edition of Quarterly J. Pure Appl. Math. in honour of John Coates.
- [15] ———, *The dichotomy between structure and randomness, arithmetic progressions, and the primes*, Proceedings of the ICM 2006, vol. 1.
- [16] R. C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A-B **285** (1977), no. 16, A981–A983.
- [17] ———, *The Hardy-Littlewood method*, Second edition. Cambridge Tracts in Mathematics, **125**. Cambridge University Press, Cambridge, 1997. xiv+232 pp

CENTRE FOR MATHEMATICAL SCIENCES, WILBERFORCE ROAD, CAMBRIDGE CB3 0WA, ENGLAND

E-mail address: b.j.green@dpmms.cam.ac.uk

UCLA DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA 90095-1596, USA

E-mail address: tao@math.ucla.edu

THE MÖBIUS FUNCTION IS STRONGLY ORTHOGONAL TO NILSEQUENCES

BEN GREEN AND TERENCE TAO

ABSTRACT. We show that the Möbius function $\mu(n)$ is strongly asymptotically orthogonal to any polynomial nilsequence $(F(g(n)\Gamma))_{n \in \mathbb{N}}$. Here, G is a simply-connected nilpotent Lie group with a discrete and cocompact subgroup Γ (so G/Γ is a *nilmanifold*), $g : \mathbb{Z} \rightarrow G$ is a polynomial sequence and $F : G/\Gamma \rightarrow \mathbb{R}$ is a Lipschitz function. More precisely, we show that $|\frac{1}{N} \sum_{n=1}^N \mu(n) F(g(n)\Gamma)| \ll_{F,G,\Gamma,A} \log^{-A} N$ for all $A > 0$. In particular, this implies the *Möbius and Nilsequence conjecture* $MN(s)$ from our earlier paper [8] for every positive integer s . This is one of two major ingredients in our programme in [8] to establish a large number of cases of the *generalised Hardy-Littlewood conjecture*, which predicts how often a collection $\psi_1, \dots, \psi_t : \mathbb{Z}^d \rightarrow \mathbb{Z}$ of linear forms all take prime values. The proof is a relatively quick application of the results in our recent companion paper [9].

We give some applications of our main theorem. We show, for example, that the Möbius function is uncorrelated with any bracket polynomial such as $n\sqrt{3}\lfloor n\sqrt{2} \rfloor$. We also obtain a result about the distribution of nilsequences $(a^n x \Gamma)_{n \in \mathbb{N}}$ as n ranges only over the primes.

1. INTRODUCTION

IMPORTANT REMARK. This paper is intimately tied to, and is intended to be read in conjunction with, the longer companion paper [9], which proves results about the distribution of finite polynomial orbits on nilmanifolds. In particular, we shall make heavy use of the notation and lemmas from that paper.

The aim of this paper is to establish what the authors have been referring to as the *Möbius and Nilsequence conjecture* $MN(s)$, first stated as [8, Conjecture 8.5]. Roughly speaking, this states that the *Möbius function* $\mu(n)$, defined as $(-1)^k$ when n is the product of k distinct primes, and 0 otherwise, is asymptotically strongly orthogonal to any Lipschitz s -step nilsequence $(F(a^n x))_{n \in \mathbb{Z}}$, in the sense that the inner product

$$\mathbb{E}_{n \in [N]} \mu(n) F(a^n x)$$

of these two functions on $[N] := \{1, \dots, N\}$ decays to zero faster than any fixed power of $1/\log N$. Here and in the sequel we use the averaging notation $\mathbb{E}_{x \in X} f(x) := \frac{1}{|X|} \sum_{x \in X} f(x)$ for any finite set X . Recall also that an *Lipschitz s -step nilsequence* is any sequence of the form $F(a^n x)$, where a is an element of an s -step connected and simply connected nilpotent Lie group G , x is an element of the *nilmanifold* G/Γ for some discrete cocompact subgroup $\Gamma \leq G$ of G , and $F : G/\Gamma \rightarrow \mathbb{R}$ is a Lipschitz function.

The difficulty of this conjecture increases with s . The case $s = 0$ of this conjecture is the estimate

$$\mathbb{E}_{n \in [N]} \mu(n) \ll_A \log^{-A} N.$$

The stronger estimate

$$\mathbb{E}_{n \in [N]} \mu(n) \ll e^{-c\sqrt{\log N}}$$

is essentially equivalent to the prime number theorem (with classical error term).

The case $s = 1$ may be reduced by Fourier analysis to the estimate

$$|\mathbb{E}_{n \in [N]} \mu(n) e(\alpha n)| \ll_A \log^{-A} N \quad (1.1)$$

where $e(x) := e^{2\pi i x}$, required to hold uniformly for all $\alpha \in \mathbb{R}$. This was established by Davenport [3] in the 1930s by modifying Vinogradov's method of bilinear forms (or "Type I and Type II sums").

In the case $s = 2$ the conjecture was established by the authors in [7]. For a more complete discussion of the conjecture and the reasons for being interested in it (and in particular, its applications to the generalised Hardy-Littlewood conjecture on the number of solutions to systems of linear equations in which the unknowns are all prime) the reader may refer to the introduction of [7], the first several sections of [8], or any of the expository articles [4, 5, 14, 15].

In this paper we settle the Möbius and Nilsequence conjecture. In fact, we shall prove the marginally stronger result that the Möbius function is asymptotically strongly orthogonal to any *polynomial* nilsequence $(F(g(n)\Gamma))_{n \in \mathbb{Z}}$.

Theorem 1.1 (Main Theorem). *Let G/Γ be a nilmanifold of some dimension $m \geq 1$, let G_\bullet be a filtration¹ of G of some degree $d \geq 1$, and let $g \in \text{poly}(\mathbb{Z}, G_\bullet)$ be a polynomial sequence². Suppose that G/Γ has a Q -rational Mal'cev basis³ \mathcal{X} for some $Q \geq 2$, defining a metric $d_{\mathcal{X}}$ on G/Γ . Suppose that $F : G/\Gamma \rightarrow [-1, 1]$ is a Lipschitz function. Then we have the bound*

$$|\mathbb{E}_{n \in [N]} \mu(n) F(g(n)\Gamma)| \ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \log^{-A} N$$

for any $A > 0$ and $N \geq 2$. The implied constant is ineffective.

Remarks. By specialising to the linear case $g(n) := a^n h$ for some $a, h \in G$ (and using the existence of Q -rational Mal'cev bases, see [9, Proposition A.9]), Theorem 1.1 immediately implies the *Möbius and nilsequences conjecture* [8, Conjecture 8.5]. In fact it gives a somewhat more precise result, since the dependence on Q and $\|F\|_{\text{Lip}}$ is given quite explicitly. For the application of Theorem 1.1 in [8], however, knowledge of these dependencies is not necessary.

The ineffectivity of the bound in Theorem 1.1 already occurs for sufficiently large A in the 1-step case (which, as mentioned before, is essentially (1.1)), and is ultimately due to the well-known ineffective bounds on Siegel zeroes. On the other hand, the remainder of the argument is effective, and so any effective bound for Siegel's theorem would imply effective bounds for Theorem 1.1. In particular, this would be the case if one assumed GRH. In fact, in that case it is not difficult to see from modifying the arguments below

¹In other words, $G_\bullet = (G_i)_{i=0}^d$ where $G = G_0 \supseteq G_1 \supseteq \dots \supseteq G_d$ is a descending sequence of Lie groups and $[G_i, G_j] \subseteq G_{i+j}$ for all $i, j \geq 0$, with the convention that G_i is trivial for $i > d$; see [9, Definition 1.2].

²A sequence $g : \mathbb{Z} \rightarrow G$ lies in $\text{poly}(\mathbb{Z}, G_\bullet)$ if $\partial_{h_1} \dots \partial_{h_i} g$ takes values in G_i for all $h_1, \dots, h_i \in \mathbb{Z}$ and $i \geq 0$, where $\partial_h g(n) := g(n+h)g(n)^{-1}$; see [9, Definition 1.11] and the ensuing discussion.

³The notion of a Q -rational Mal'cev basis is defined in [9, Definition 2.6] and the construction of the metric $d_{\mathcal{X}}$ is given in the same section.

that we can replace the logarithmic decay $\log^{-A} N$ by polynomial decay N^{-c} for some $c > 0$ depending only on d and m .

The authors learnt in [9] that it is in many ways more natural to consider the class of polynomial sequences $\text{poly}(\mathbb{Z}, G_\bullet)$ rather than simply the class of linear sequences $n \mapsto a^n x$. This is ultimately due to the stability of the polynomial class under a wide variety of operations, such as pointwise multiplication. On the other hand, these two categories are certainly closely related (and are, in some sense, equivalent): see [12] for further discussion.

ACKNOWLEDGEMENTS. The first author is partly supported by a Leverhulme Prize. The second author is supported by a grant from the Macarthur Foundation and by NSF grant DMS-0649473.

2. REDUCING TO THE EQUIDISTRIBUTED CASE

To prove Theorem 1.1, we will apply [9, Theorem 1.19] to decompose g as a product $\varepsilon g' \gamma$ where ε is “smooth”, γ is “rational” and g' is highly equidistributed in some closed subgroup $G' \subseteq G$. We will recall the precise statement shortly.

In this section we shall show how the rather harmless factors ε and γ in the above factorisation may be eliminated, and then make an additional reduction to the case $\int_{G/\Gamma} F = 0$ (using the Haar measure on G/Γ , of course). This leaves us with the task of proving an “equidistributed” case of Theorem 1.1: see Proposition 2.1 below.

For the rest of the paper, all constants c, C , including those in the asymptotic notation \ll and $O()$, are allowed to depend on m and d . Different occurrences of the letters c, C may represent different constants; typically we will have $0 < c \ll 1 \ll C < \infty$. For ease of notation we drop the subscript whenever Lipschitz norms are mentioned, so $\|F\|_{\text{Lip}}$ becomes simply $\|F\|$.

Recall from [9, Definition 1.3(v)] that a sequence $(g(n)\Gamma)_{n \in [N]}$ in a nilmanifold is *totally δ -equidistributed* if we have

$$|\mathbb{E}_{n \in P} F(g(n)\Gamma)| \leq \delta \|F\| \quad (2.1)$$

for all Lipschitz functions $F : G/\Gamma \rightarrow \mathbb{C}$ with $\int_{G/\Gamma} F = 0$ and all arithmetic progressions $P \subseteq [N]$ of length at least δN .

In the next section we shall establish the following result about the lack of correlation of Möbius with equidistributed nilsequences.

Proposition 2.1 (Möbius is orthogonal to equidistributed sequences). *Let $m \geq 0$, $d \geq 1$ be integers and let $N \geq 1$ be an integer parameter which is sufficiently large depending on m and d . Let δ , $0 < \delta < 1/2$, and $Q \geq 2$ be real parameters. Let G/Γ be an m -dimensional nilmanifold, and suppose that G_\bullet is a filtration of degree d . Suppose that G/Γ has a Q -rational Mal'cev basis \mathcal{X} adapted to the filtration G_\bullet . Let $g \in \text{poly}(\mathbb{Z}, G_\bullet)$ and suppose that $(g(n)\Gamma)_{n \in [N]}$ is totally δ -equidistributed. Then for any function $F : G/\Gamma \rightarrow \mathbb{R}$ with $\int_{G/\Gamma} F = 0$ and for any arithmetic progression $P \subseteq [N]$ of size at least N/Q , we have the bound*

$$|\mathbb{E}_{n \in [N]} \mu(n) 1_P(n) F(g(n)\Gamma)| \ll \delta^c Q \|F\| \log N.$$

The proof of Proposition 2.1 proceeds via the method of Type I/II sums, which is also known as the method of bilinear forms. This is the same method that one might use to tackle the “minor arcs” case of (1.1), where α is not close to a rational with small denominator. We will describe it in detail in the next section. Our task for the remainder of this section is to reduce Theorem 1.1 to Proposition 2.1.

Proof that Proposition 2.1 implies Theorem 1.1. We start with a brief overview. The main ingredient of this argument is [9, Theorem 1.19], that is to say the factorization $g = \varepsilon g' \gamma$ mentioned above. In addition to that we require estimates for sums of the type $\mathbb{E}_{n \in [N]} \mu(n) 1_P(n)$, where $P \subseteq [N]$ is a progression. After standard harmonic analysis, such bounds ultimately depend on results about the zeros of L -functions $L(s, \chi)$, and as such this is analysis of the same type as would be used to establish the “major arc” cases of (1.1). Finally, a fair amount of what might be called “quantitative nil-linear algebra” is required to keep track of the various nilmanifolds and Lipschitz functions involved in the argument. Here we draw repeatedly on the material assembled in [9, Appendix A] for this purpose; we encourage the reader to gloss over these essentially routine issues on a first reading.

We now turn to the details. We allow all implied constants to depend on m and d .

Let the hypotheses be as in Theorem 1.1. To simplify the notation slightly we will also assume that $\|F\| \geq 1$; the case $\|F\| < 1$ can easily be deduced from that case. By dividing out by $\|F\|$ we may in fact normalize and assume that $\|F\| = 1$.

We may of course take $A \geq 1$. We may also assume that $Q \leq \log N$, since the claim is vacuously true otherwise; thus \mathcal{X} is now a log N -rational Mal’cev basis. By increasing A if necessary, it will suffice to show an estimate of the form

$$|\mathbb{E}_{n \in [N]} \mu(n) F(g(n) \Gamma)| \ll_A \log^{-A+O(1)} N. \quad (2.2)$$

Let B be a parameter (depending on A) to be specified later. We may assume that N is sufficiently large depending on A, B . By [9, Theorem 1.19] (with $M_0 := \log N$) we can find an integer M ,

$$\log N \leq M \ll \log^{O_B(1)} N,$$

a rational subgroup $G' \subseteq G$, a Mal’cev basis \mathcal{X}' for G'/Γ' (where $\Gamma' := G \cap \Gamma$) in which each element is an M -rational combination (see [9, Definition 1.21]) of the elements of \mathcal{X} , and a decomposition

$$g = \varepsilon g' \gamma \quad (2.3)$$

into polynomial sequences $\varepsilon, g', \gamma \in \text{poly}(\mathbb{Z}, G_\bullet)$ with the following properties:

- (i) $\varepsilon : \mathbb{Z} \rightarrow G_\bullet$ is (M, N) -smooth (see [9, Definition 1.22] for a definition);
- (ii) $g' : \mathbb{Z} \rightarrow G'$ takes values in G' , and the finite sequence $(g'(n) \Gamma')_{n \in [N]}$ is totally M^{-B} -equidistributed in G'/Γ' , using the metric $d_{\mathcal{X}'}$ on G'/Γ' ;
- (iii) $\gamma : \mathbb{Z} \rightarrow G$ is M -rational (see [9, Definition 1.21]), and $(\gamma(n) \Gamma)_{n \in \mathbb{Z}}$ is periodic with period $1 \leq q \leq M$.

From (2.3) we have

$$\mathbb{E}_{n \in [N]} \mu(n) F(g(n) \Gamma) = \mathbb{E}_{n \in [N]} \mu(n) F(\varepsilon(n) g'(n) \gamma(n) \Gamma). \quad (2.4)$$

The sequence $(\gamma(n) \Gamma)_{n \in \mathbb{Z}}$ is periodic with some period q , $1 \leq q \leq M$. For each $j = 0, 1, \dots, q-1$ let $\gamma_j := \{\gamma(j)\}$ be the fractional part of $\gamma(j)$ with respect to Γ , thus

$\gamma_j \Gamma = \gamma(j) \Gamma$ and all the coordinates $\psi_{\mathcal{X}}(\gamma_j)$ lie in $[0, 1)$. This construction is described in [9, Lemma A.14].

Now by [9, Lemma A.12], the coordinates $\psi_{\mathcal{X}}(\gamma(j))$ lie in $\frac{1}{M'} \mathbb{Z}^m$ for some $M' \ll M^{O(1)}$. Since $\gamma_j = \gamma(j) \eta$ for some η with integer coordinates, it follows from [9, Lemma A.3] that the coordinates $\psi_{\mathcal{X}}(\gamma_j)$ are rationals with height $\ll M^{O(1)}$.

We now take advantage of the periodicity of $\gamma(n) \Gamma$ to split the right-hand side of (2.4) as

$$\sum_{j=0}^{q-1} \mathbb{E}_{n \in [N]} \mu(n) 1_{n \equiv j \pmod{q}} F(\varepsilon(n) g'(n) \gamma_j \Gamma); \quad (2.5)$$

By the right-invariance of d , the (M, N) -smoothness of ε (see [9, Definition 1.21]) and the 1-Lipschitz bound on F we see that

$$\begin{aligned} |F(\varepsilon(n) g'(n) \gamma_j \Gamma) - F(\varepsilon(n_0) g'(n) \gamma_j \Gamma)| &\leq d_{\mathcal{X}}(\varepsilon(n) g'(n) \gamma_j, \varepsilon(n_0) g'(n) \gamma_j) \\ &= d_{\mathcal{X}}(\varepsilon(n_0), \varepsilon(n)) \\ &\leq \log^{-A} N. \end{aligned}$$

whenever if $|n - n_0| \leq \frac{N}{M \log^A N}$. Hence if we split each progression $n \equiv j \pmod{q}$ into further progressions $P_{j,k}$ for $k = O(M \log^A N)$, each having diameter at most $\frac{N}{M \log^A N}$, we see that (2.5) is equal to

$$\sum_{j,k} \mathbb{E}_{n \in [N]} \mu(n) 1_{P_{j,k}}(n) F(a_{j,k} g'(n) \gamma_j \Gamma) + O(\log^{-A} N). \quad (2.6)$$

Here each $a_{j,k} := \varepsilon(n_{0,j,k})$ for some $n_{0,j,k} \in P_{j,k}$; by the definition of what it means for $\varepsilon : \mathbb{Z} \rightarrow G$ to be (M, N) -smooth (i.e. [9, Definition 1.21]), it follows that $d_{\mathcal{X}}(a_{j,k}, \text{id}_G) \leq M$ and hence, by [9, Lemma A.4], that

$$|\psi_{\mathcal{X}}(a_{j,k})| \ll M^{O(1)}. \quad (2.7)$$

If N is sufficiently large depending on A and B then $N \geq 10M \log^A N$ (say), and this partition of $[N]$ may be arranged in such a way that

$$|P_{j,k}| \geq \frac{N}{2qM \log^A N} \geq \frac{N}{2M^2 \log^A N}.$$

Since the number of j is at most M , and the number of k is at most $M \log^A N$, we thus see that to show (2.2) it suffices by the triangle inequality to show that

$$|\mathbb{E}_{n \in [N]} \mu(n) 1_{P_{j,k}}(n) F(a_{j,k} g'(n) \gamma_j \Gamma)| \ll_A M^{-2} \log^{-2A+O(1)} N \quad (2.8)$$

for each j, k .

Fix j, k . Write $H_j := \gamma_j^{-1} G' \gamma_j$ and let $g_j : \mathbb{Z} \rightarrow H_j$ be the sequence defined by $g_j(n) := \gamma_j^{-1} g'(n) \gamma_j$. It is clear that each g_j is a polynomial sequence with coefficients in the filtration $(H_j)_{\bullet} := \gamma_j^{-1} G'_{\bullet} \gamma_j$.

Set $\Lambda_j := \Gamma \cap H_j$ and define functions

$$F_{j,k} : H_j / \Lambda_j \rightarrow [-1, 1]$$

by the formula

$$F_{j,k}(x \Lambda_j) := F(a_{j,k} \gamma_j x \Gamma).$$

Then (2.8) can be rewritten as

$$|\mathbb{E}_{n \in [N]} \mu(n) 1_{P_{j,k}}(n) F_{j,k}(g_j(n) \Lambda_j)| \ll_A M^{-2} \log^{-2A+O(1)} N. \quad (2.9)$$

Suppose for the moment that $F_{j,k}$ were a constant function. Recall that $P_{j,k}$ has common difference $q \leq M$. We may thus apply Proposition A.2 (with A replaced by a sufficiently large exponent A' depending on A and B) to obtain the desired claim, since $M \ll \log^{O_B(1)} N$. Therefore we may subtract off the mean of $F_{j,k}$ and assume without loss of generality that $\int_{H_j/\Lambda_j} F_{j,k} = 0$. This may cause $F_{j,k}$ to take values in $[-2, 2]$ rather than $[-1, 1]$, but we can easily counter this trivial issue by dividing $F_{j,k}$ by two.

In a moment we shall use Proposition 2.1 to estimate the terms appearing here. Before doing that we record quantitative rationality properties of the nilmanifold H_j/Λ_j , as well as a Lipschitz bound on $\|F_{j,k}\|$.

Claim. There is a Mal'cev basis \mathcal{Y}_j for H_j/Λ_j adapted to the filtration $(H_j)_\bullet$ such that each \mathcal{Y}_j is an M^C -rational combination of the X_i . With respect to the metric $d_{\mathcal{Y}_j}$ on H_j/Λ_j induced by this basis, the polynomial sequence $g_j \in \text{poly}(\mathbb{Z}, (H_j)_\bullet)$ is $M^{-cB+O(1)}$ -totally equidistributed for some $c > 0$ depending only on m, d , and we have $\|F_{j,k}\| \leq M^{O(1)}$.

Proof. We shall apply suitable combinations of the lemmas in [9, Appendix A]. The existence of \mathcal{Y}_j follows from Proposition A.9 and Lemma A.13 of [9] together with the fact that each γ_j has rational coordinates with height $M^{O(1)}$. Now the map $x \mapsto F(a_{j,k} \gamma_j x \Gamma)$ on G/Γ has Lipschitz constant at most $M^{O(1)}$ by [9, Lemma A.5] and the bounds $|\psi_{\mathcal{X}}(a_{j,k})|, |\psi_{\mathcal{X}}(\gamma_j)| \leq M^{O(1)}$. The last statement of the claim follows from [9, Lemma A.17]. Finally, the statement about the quantitative equidistribution of g_j (in which is hidden a slight technical subtlety) is addressed by Lemma B.1 in the appendix of this paper. \square

Let us now apply Proposition 2.1 to (2.9). We apply the proposition with parameters (which we distinguish using tildes) as follows: $\tilde{G} := H_j$, $\tilde{\Gamma} := \Lambda_j$, $\tilde{G}_\bullet := (H_j)_\bullet$, $\tilde{g} := g_j$, $\tilde{X} := \mathcal{Y}_j$, $\tilde{Q} := M^{O(1)}$, $\tilde{F} := F_{j,k}$ and $\tilde{\delta} := M^{-cB+O(1)}$. We quickly see that (2.9) is bounded by $O\left(M^{-cB+O(1)} \log^{O(A)} N\right)$. Choosing B sufficiently large depending on A , we obtain (2.9) as claimed. \square

3. THE EQUIDISTRIBUTED CASE: TYPE I AND II SUMS

In this section we establish Proposition 2.1 using Vinogradov's method of Type I and II sums in the form due to Vaughan [16]. More precisely, we will use the following proposition.

Proposition 3.1 (Method of Type I/II sums). *Let $f : \mathbb{N} \rightarrow \mathbb{C}$ be a function with $\|f\|_\infty \leq 1$ such that*

$$|\mathbb{E}_{N < n \leq 2N} \mu(n) \overline{f(n)}| \geq \varepsilon$$

for some $\varepsilon > 0$. Then one of the following statements holds:

- (Type I sum is large) *There exists an integer $1 \leq K \leq N^{2/3}$ such that*

$$|\mathbb{E}_{N/k < w \leq 2N/k} f(kw)| \gg (\varepsilon / \log N)^{O(1)} \quad (3.1)$$

for $\gg (\varepsilon/\log N)^{O(1)}K$ integers k such that $K < k \leq 2K$.

- (Type II sum is large) *There exist integers K, W with $\frac{1}{2}N^{1/3} \leq K \leq 4N^{2/3}$ and $N/4 \leq KW \leq 4N$, such that*

$$|\mathbb{E}_{K < k, k' \leq 2K} \mathbb{E}_{W \leq w, w' < 2W} f(kw) \overline{f(k'w)} \overline{f(kw')} f(k'w')| \gg (\varepsilon/\log N)^{O(1)}. \quad (3.2)$$

Proof. This is [7, Proposition 4.2], specialised to the case $U = V = N^{1/3}$, and with certain explicit exponents replaced by unspecified constants $O(1)$. \square

We now begin the proof of Proposition 2.1. As before we may normalise so that $\|F\| = 1$. From this and the mean zero assumption, we see in particular that

$$|F(x)| \leq \text{diam}(G/\Gamma) \ll Q^{O(1)} \quad (3.3)$$

for all $x \in G/\Gamma$ (the diameter bound here is [9, Lemma A.16]).

If $\delta \leq 1/N$ then by (2.1) we have $|F(g(n)\Gamma)| \leq \delta$ for all $n \in [N]$, and the claim is trivial, so we may assume that $\delta > 1/N$. By increasing δ if necessary (and shrinking c) we thus see that we may assume that

$$\delta > N^{-\sigma} \quad (3.4)$$

for any fixed small constant $\sigma > 0$ depending only on m, d .

The basic idea, which will become clearer upon reading the details, is to make good use of the fact that one may test the quantitative equidistribution properties of a polynomial nilsequence on G/Γ by passing to the abelianisation $(G/\Gamma)_{\text{ab}}$, a phenomenon referred to in [9, Theorem 2.9] as the “quantitative Leibman Dichotomy” (cf. [12]). The abelian issues that one must then deal with are of a very similar nature to those involved in dealing with exponential sums such as $\mathbb{E}_{n \in [N]} \mu(n) e(p(n))$, where $p : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$ is an ordinary polynomial. Rather than quote results from the existing literature on this problem it is easier for us to invoke various lemmas from [9], which were stated and proved in a language which is helpful for the present paper.

Let $\varepsilon := \delta^{c_1} Q \log N$, for a constant c_1 to be specified later. We may assume that $\varepsilon < 1$, otherwise the claim is trivial from (3.3) and the triangle inequality. In particular, we have

$$Q, \log N \leq \delta^{-c_1}$$

and we will use these estimates frequently in the sequel to absorb any polynomial factors in Q or $\log N$ into a power of δ^{-c_1} .

Suppose for contradiction that Proposition 2.1 failed for these parameters. We then apply Proposition 3.1 with $f(n) := 1_P(n)F(g(n)\Gamma)$ and ε as above, concluding that either (3.1) or (3.2) holds. We deal with these two cases in turn.

The Type I case. Suppose that (3.1) holds. Thus there are $\gg \delta^{O(c_1)}K$ values of $k \in (K, 2K]$ such that

$$|\mathbb{E}_{N/k < w \leq 2N/k} 1_P(kw) F(g(kw)\Gamma)| \gg \delta^{O(c_1)}.$$

Let l denote the common difference of P ; since $|P| \geq N/Q$, we must have $1 \leq l \leq Q$. Splitting into progressions with common difference l , we see that for some $b \pmod{l}$ and

for $\gg \delta^{O(c_1)} K$ values of $k \in (K, 2K]$ we have

$$\left| \sum_{\substack{N/k < w \leq 2N/k \\ w \equiv b \pmod{l}}} 1_P(kw) F(g(kw)\Gamma) \right| \gg \delta^{O(c_1)} \frac{N}{kl}.$$

Setting $w = b + lw'$, this may be rewritten as

$$\left| \sum_{w' \in I_k} F(g(k(b + lw')\Gamma) \right| \gg \delta^{O(c_1)} \frac{N}{kl}, \quad (3.5)$$

where $I_k \subseteq [\frac{N}{2kl} - 1, \frac{N}{kl}]$ is an interval.

For each value of k for which this holds, consider the sequence $g_k : \mathbb{Z} \rightarrow G$ defined by $g_k(n) := g(kn)$ and also the sequence $\tilde{g}_k : \mathbb{Z} \rightarrow G$ defined by $\tilde{g}_k(n) = g(k(b + ln))$. It follows from [9, Corollary 6.8] that $g_k, \tilde{g}_k \in \text{poly}(\mathbb{Z}, G_\bullet)$. Now (3.5) implies that $(\tilde{g}_k(n)\Gamma)_{n \in [N_k]}$ fails to be $\delta^{O(c_1)}$ -equidistributed in G/Γ , where $N_k \sim N/kl$.

It follows from [9, Theorem 2.9] that there is a nontrivial horizontal character $\psi_k : G \rightarrow \mathbb{R}/\mathbb{Z}$ (i.e. a continuous homomorphism from G to \mathbb{R}/\mathbb{Z} which annihilates Γ) with magnitude $|\psi_k| \ll \delta^{-O(c_1)}$ such that

$$\|\psi_k \circ \tilde{g}_k\|_{C^\infty[N_k]} \ll \delta^{-O(c_1)}.$$

Recall from [9, Definition 2.10] that the $C^\infty[N]$ -norm of a polynomial $p : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ expanded in binomial coefficients as

$$p(n) = \alpha_0 + \alpha_1 \binom{n}{1} + \cdots + \alpha_d \binom{n}{d}, \quad (3.6)$$

is defined by

$$\|p\|_{C^\infty[N]} := \sup_{1 \leq j \leq d} N^j \|\alpha_j\|_{\mathbb{R}/\mathbb{Z}}.$$

By [9, Lemma 8.4] (specialised to the single-parameter case $t = 1$), there is some $q_k \ll \delta^{-O(c_1)}$ such that

$$\|q_k \psi_k \circ g_k\|_{C^\infty[N_k]} \ll \delta^{-O(c_1)}.$$

Pigeonholing in the possible choices of $q_k \psi_k$, we may find some ψ with $0 < |\psi| \ll \delta^{-O(c_1)}$ such that

$$\|\psi \circ g_k\|_{C^\infty[N_k]} \ll \delta^{-O(c_1)} \quad (3.7)$$

for $\gg \delta^{O(c_1)} K$ values of $k \in (K, 2K]$.

Write

$$\psi \circ g(n) = \beta_d n^d + \cdots + \beta_0. \quad (3.8)$$

Then

$$\psi \circ g_k(n) = \beta_d k^d n^d + \cdots + \beta_0. \quad (3.9)$$

We would like to use this and (3.7) to conclude that the coefficients $k^j \beta_j$ are close to being integer (or rational with small denominator). This will follow from a simple lemma.

Lemma 3.2. *Suppose that $p : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a polynomial of the form $p(n) = \beta_d n^d + \cdots + \beta_0$. Then there is some $q \geq 1$, $q = O(1)$, such that $\|q\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll N^{-j} \|p\|_{C^\infty[N]}$ for $j = 1, \dots, d$.*

Proof. Consider the representation (3.6) which is used to define the $C^\infty[N]$ -norm. Observing that β_j can be written as a linear combination of $\alpha_j, \dots, \alpha_d$ with rational coefficients of height $O(1)$, the result follows upon clearing denominators. \square

From (3.7), (3.9) and Lemma 3.2 we see that there is some $q \geq 1$, $q = O(1)$, such that

$$\|qk^j\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}(N/K)^{-j} \quad (3.10)$$

for $j = 1, 2, \dots, d$ and for at least $\delta^{O(c_1)}K$ values of $k \in (K, 2K]$.

Fix j , $1 \leq j \leq d$. To pass from the j^{th} powers k^j to more general integers we shall need the following Waring-type result.

Lemma 3.3. *Let $K \geq 1$ be an integer, and suppose that $S \subseteq [K]$ is a set of size αK . Suppose that $t \geq 2^j + 1$. Then $\gg_{j,t} \alpha^{2t} K^j$ integers in the interval $[tK^j]$ can be written in the form $k_1^j + \dots + k_t^j$, $k_1, \dots, k_t \in S$.*

Proof. It is a well-known consequence of Hardy and Littlewood's asymptotic formula for Waring's problem (see e.g. [17]) that the number of solutions to

$$x_1^j + \dots + x_t^j = M, \quad x_1, \dots, x_t \in [K]$$

is $\ll_{j,t} K^{t-j}$ uniformly in M provided that $t \geq 2^j + 1$. (In fact, by subsequent work, such a result is known for much smaller values of t when j is large.) Let $X = \{k^j : k \in S\}$ and let $r(n)$ be the number of representations of n as the sum of t elements of X . Then by the Cauchy-Schwarz inequality and the preceding remarks we have

$$\alpha^{2t} K^{2t} = \left(\sum_n r(n) \right)^2 \leq |tX| \sum_n r(n)^2 \ll_j |tX| K^{2t-j},$$

which implies the result. \square

By (3.10) and Lemma 3.3 it follows that

$$\|ql\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}(K/N)^j$$

for $\gg \delta^{O(c_1)}K^j$ values of $l \in [10^d K^j]$.

The following lemma, which is [9, Lemma 3.2], may be applied to this situation.

Lemma 3.4 (Strongly recurrent linear functions are highly non-diophantine). *Let $\alpha \in \mathbb{R}$, $0 < \sigma < 1/2$, and $0 < \mu \leq \sigma/2$, and let $I \subseteq \mathbb{R}/\mathbb{Z}$ be an interval of length μ such that $\alpha n \in I$ for at least σN values of $n \in [N]$. Then there is some $k \in \mathbb{Z}$ with $0 < |k| \ll \sigma^{-O(1)}$ such that $\|k\alpha\|_{\mathbb{R}/\mathbb{Z}} \ll \mu \sigma^{-O(1)}/N$.* \square

Let us attempt to apply this lemma with $\sigma \gg \delta^{O(c_1)}$ and $\mu \ll \delta^{-O(c_1)}(K/N)^j$. If N is sufficiently large and the exponent σ in (3.4) is sufficiently small, we see using the bound $K/N \leq N^{-1/3}$ that the hypotheses of the lemma are satisfied and that such an application is permissible. The conclusion is that there is some q' , $1 \leq q' \ll \delta^{-O(c_1)}$, such that

$$\|qq'\beta_i\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}N^{-i}. \quad (3.11)$$

Writing $\tilde{\psi} := qq'\psi$, it follows from (3.8) and (3.11) that for any n we have the bound

$$\|\tilde{\psi} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)}n/N.$$

If $N' := \delta^{C_1} N$ for some sufficiently large C , and if $n \in [N']$, this implies that

$$\|\tilde{\psi} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \leq 1/10. \quad (3.12)$$

Now set $\tilde{F} : G/\Gamma \rightarrow [-1, 1]$ to be the function $\tilde{F} := \eta \circ \tilde{\psi}$, where $\eta : \mathbb{R}/\mathbb{Z} \rightarrow [-1, 1]$ is a function of Lipschitz norm $O(1)$ and mean zero which equals 1 on $[-1/10, 1/10]$. Then we have $\int_{G/\Gamma} \tilde{F} = 0$ and $\|\tilde{F}\| \ll \delta^{-O(c_1)}$. From (3.12), we have

$$|\mathbb{E}_{n \in [N']} \tilde{F}(g(n)\Gamma)| \geq 1 > \delta \|\tilde{F}\|,$$

provided that c_1 is chosen sufficiently small. This is contrary to the assumption that $(g(n)\Gamma)_{n \in [N]}$ is δ -totally equidistributed.

The Type II case. This is in many ways very closely similar to the Type I case, as the reader will see. Recall the situation that (3.2) puts us in (with our choice of ε): there are K, W with $\frac{1}{2}N^{1/3} \leq K \leq 4N^{2/3}$ and $N/4 \leq KW \leq 4N$ such that

$$|\mathbb{E}_{K < k, k' \leq 2K} \mathbb{E}_{W < w, w' \leq 2W} f(kw)f(kw')f(k'w)f(k'w')| \gg \delta^{O(c_1)},$$

where $f(n) = 1_P(n)F(g(n)\Gamma)$. Writing the left-hand side here as

$$\mathbb{E}_{K < k, k' \leq 2K} |\mathbb{E}_{W < w \leq 2W} f(kw)f(k'w)|^2,$$

we see that there are $\gg \delta^{O(c_1)} K^2$ pairs $(k, k') \in (K, 2K]^2$ such that

$$|\mathbb{E}_{W < w \leq 2W} f(kw)f(k'w)| \gg \delta^{O(c_1)}.$$

Written out in full, for each such pair (k, k') we have

$$|\mathbb{E}_{W < w \leq 2W} 1_P(kw)1_P(k'w)F(g(kw)\Gamma)F(g(k'w)\Gamma)| \gg (\varepsilon/\log N)^{O(1)}.$$

Writing l for the common difference of P (thus $1 \leq l \leq Q$) we see that there is some $b \pmod l$ such that for $\gg (\varepsilon/\log N)^{O(1)} K^2$ pairs (k, k') we have

$$\sum_{\substack{W < w \leq 2W \\ w \equiv b \pmod l}} 1_P(kw)1_P(k'w)F(g(kw)\Gamma)F(g(k'w)\Gamma) \gg \delta^{O(c_1)} \frac{W}{l}.$$

Setting $w = lw' + b$, this may be written as

$$|\sum_{w' \in I_{k,k'}} F(g(k(b+lw'))\Gamma)F(g(k'(b+lw'))\Gamma)| \gg \delta^{O(c_1)} \frac{W}{l}, \quad (3.13)$$

where $I_{k,k'} \subseteq (\frac{W}{l} - 1, \frac{2W}{l}]$ is an interval. Since $1 \leq l \leq Q$, which is bounded by a small power of N , and $W \gg N^{1/3}$, this is contained in $[\frac{W}{2l}, \frac{2W}{l}]$.

For each k, k' for which this holds, consider the sequence $g_{k,k'} : \mathbb{Z} \rightarrow G \times G$ defined by $g_{k,k'}(n) = (g(kn), g(k'n))$, and also the sequence $\tilde{g}_{k,k'} : \mathbb{Z} \rightarrow G \times G$ defined by $\tilde{g}_{k,k'}(n) = (g(k(b+ln)), g(k'(b+ln)))$. It follows from [9, Corollary 6.8] that $g_{k,k'}, \tilde{g}_{k,k'} \in \text{poly}(\mathbb{Z}, G_\bullet \times G_\bullet)$. Now from (3.13) we see that the sequence $(\tilde{g}_{k,k'}(n)(\Gamma \times \Gamma))_{n \in [N_{k,k'}]}$ fails to be $\delta^{O(c_1)}$ -equidistributed in $(G/\Gamma) \times (G/\Gamma)$, for some $N_{k,k'} \in [\frac{W}{2l}, \frac{2W}{l}]$.

It follows from [9, Theorem 2.9] that there is a nontrivial horizontal character $\psi_{k,k'} : G \times G \rightarrow \mathbb{R}/\mathbb{Z}$ with $|\psi_k| \ll \delta^{-O(c_1)}$ such that

$$\|\psi_{k,k'} \circ \tilde{g}_{k,k'}\|_{C^\infty[N_{k,k'}]} \ll \delta^{-O(c_1)}.$$

By [9, Lemma 8.4] there is some $q_{k,k'} \ll \delta^{-O(c_1)}$ such that

$$\|q_{k,k'} \psi_{k,k'} \circ g_{k,k'}\|_{C^\infty[N_{k,k'}]} \ll \delta^{-O(c_1)}.$$

Pigeonholing in the possible choices of $q_{k,k'} \psi_{k,k'}$, we may find some ψ with $0 < |\psi| \ll \delta^{-O(c_1)}$ such that

$$\|\psi \circ g_{k,k'}\|_{C^\infty[N_{k,k'}]} \ll \delta^{-O(c_1)} \quad (3.14)$$

for $\gg \delta^{O(c_1)} K^2$ pairs $k, k' \in (K, 2K]$.

Write $\psi = \psi_1 \oplus \psi_2$, where $\psi_1, \psi_2 : G \rightarrow \mathbb{R}/\mathbb{Z}$ are horizontal characters, not both zero. If

$$\psi_1 \circ g(n) = \beta_d n^d + \cdots + \beta_0$$

and

$$\psi_2 \circ g(n) = \beta'_d n^d + \cdots + \beta'_0$$

then

$$\psi \circ g_{k,k'}(n) = (\beta_d k^d + \beta'_d k'^d) n^d + \cdots + (\beta_0 + \beta'_0),$$

By Lemma 3.2 and (3.14) there is some $1 \leq q \ll \delta^{-O(c_1)}$ such that

$$\|q(k^j \beta_j + k'^j \beta'_j)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)} N_{k,k'}^{-j} \ll \delta^{-O(c_1)} (K/N)^j$$

for $j = 1, 2, \dots, d$ and for $\gg \delta^{O(c_1)} K^2$ pairs $k, k' \in (K, 2K]$.

Suppose, without loss of generality, that $\psi_1 \neq 0$. Selecting some k' that occurs in $\gg \delta^{O(c_1)} K$ of the pairs k, k' and subtracting, we see that

$$\|q k^j \beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(c_1)} (K/N)^j \quad (3.15)$$

for $\gg \delta^{O(c_1)} K$ values of $k \in (-K, K)$. Using the bounds $K \gg N^{1/3}$ and (3.4) it follows that we may ignore the contribution of $k = 0$, that is to say (3.15) holds for $\gg \delta^{O(c_1)} K$ values of $k \in [1, K]$.

Remark. Note carefully that (3.15) carries no information when $k = 0$. In our treatment of Type I sums there was no need for a lower bound on K , but such an assumption is essential if one has any desire to bound Type II sums.

The estimate (3.15) is identical to (3.10). We may now repeat the arguments used to obtain a contradiction to (3.10) in Type I case. The proof of Proposition 2.1 and thus Theorem 1.1 is now complete. \square

The main business of the paper is now complete. In the next section we give a brief discussion of how our argument compares with the classical Hardy-Littlewood method. After that we give a number of applications of Theorem 1.1.

4. REMARKS ON A NILPOTENT HARDY-LITTLEWOOD METHOD

It may be of interest to interpret our method in terms of the “major and minor arcs” terminology of the Hardy-Littlewood method. Recall that to prove Davenport’s estimate

$$|\mathbb{E}_{n \in [N]} \mu(n) e(\alpha n)| \ll_A \log^{-A} N$$

one divides into two cases: the *major arcs* where α is close to a rational with small denominator, and the *minor arcs* where it is not. The major arcs are handled using

L -function technology as in Appendix A, and the minor arcs are handled using Type I/II sums as in Proposition 3.1.

Suppose that we are considering the sum

$$\mathbb{E}_{n \in [N]} \mu(n) F(g(n)\Gamma),$$

where $\int_{G/\Gamma} F = 0$. Decompose g as a product $\varepsilon g' \gamma$ where ε is smooth, γ is rational and g' is highly equidistributed on some subgroup G' . Then one might think of g as a “major arc” nilsequence if $G' = \{\text{id}_G\}$, and as “minor arc” if G' is nontrivial.

To justify this terminology, observe that one may interpret $e(\alpha n)$ as $F(g(n)\Gamma)$, where $G/\Gamma = \mathbb{R}/\mathbb{Z}$, $g : \mathbb{Z} \rightarrow \mathbb{R}$ is the polynomial sequence $g(n) = \alpha n$ and the Lipschitz function F , taking values in the unit ball of the complex plane, is simply $e(\theta)$.

If $\alpha = \frac{a}{q} + \varepsilon$, where ε is small, then the decomposition $g = \varepsilon g' \gamma$ will be given by $\varepsilon(n) = \varepsilon n$, $g'(n) = \text{id}_G$ and $\gamma(n) = an/q$ and so this does indeed correspond to a “major arc nilsequence”.

If α is not close to a rational with small denominator then $g(n)$ will already be highly equidistributed on \mathbb{R}/\mathbb{Z} , and so the decomposition $g = \varepsilon g' \gamma$ has $\varepsilon = \gamma = \text{id}_G$ and $g' = g$. Thus $G' = \mathbb{R}$ is nontrivial and this corresponds to a “minor arc nilsequence”.

5. ON BRACKET POLYNOMIALS

By a *bracket polynomial* we mean an object formed from the scalar field \mathbb{R} and the indeterminate n using finitely many instances of the standard arithmetic operations $+$, \times together with the integer part operation $\lfloor \cdot \rfloor$ and the fractional part operation $\{ \cdot \}$. The following are all bracket polynomials: $n^2 + n\sqrt{2}$, $n\sqrt{2} \lfloor n\sqrt{3} \rfloor$ and $\{n^3\sqrt{2} + n^7 \lfloor n\sqrt{5} \rfloor + \sqrt{7}\}$. One may associate a notion of *complexity* to any bracket polynomial $p(n)$, this being (for instance) the least number of operations $+$, \times , $\lfloor \cdot \rfloor$, $\{ \cdot \}$ required to write down p . In view of the relation $\{x\} + \lfloor x \rfloor = x$, it is not strictly speaking necessary to retain both the integer and fractional part operations, but we do so here for convenience. Dispensing with one of them would slightly alter the definition of complexity.

The following remarkable theorem of Bergelson and Leibman [2] demonstrates a close link between bracket polynomials and nilmanifolds. If G/Γ is a nilmanifold with Mal'cev basis \mathcal{X} then recall from [9, Lemma A.14] that the coordinate map $\psi : G \rightarrow \mathbb{R}^m$ provides an identification between G/Γ and $[0, 1)^m$. Write τ_1, \dots, τ_m for the individual coordinate maps from G/Γ to $[0, 1)$, that is to say τ_i is the composition of ψ with the map $(t_1, \dots, t_m) \mapsto t_i$.

Theorem 5.1 (Bergelson-Leibman). *The functions of the form $n \mapsto \{p(n)\}$, where p is a bracket polynomial, coincide with the functions of the form $n \mapsto \tau_i(g(n)\Gamma)$, where G/Γ is a nilmanifold equipped with a Mal'cev basis \mathcal{X} and $g : \mathbb{Z} \rightarrow G$ is a polynomial map with coefficients in some filtration G_\bullet . The rationality of \mathcal{X} , the dimension of G , the degree of g and the rationality of G_\bullet may all be bounded in terms of the complexity of p , and conversely the complexity of p may be bounded in terms of these quantities.*

In fact, Bergelson and Leibman prove a number of rather refined variants of this type of result, and they also give a comprehensive and edifying discussion of bracket polynomials in general. At first glance it appears that one might immediately combine

Theorem 5.1 with Theorem 1.1 to obtain a result about the correlation of the Möbius function with bracket polynomials. There is a serious catch, however: the coordinate functions τ_i are not continuous on the nilmanifold G/Γ . Furthermore, as observed by Bergelson and Leibman, there are bracket polynomials which *cannot* be written in the form $F(g(n)\Gamma)$ for a continuous F . Indeed the results of Leibman [12] on the distribution of $(g(n)\Gamma)_{n \in \mathbb{Z}}$ imply that the sequence $(F(g(n)\Gamma))_{n \in \mathbb{Z}}$ cannot have isolated values, yet there are bracket polynomials which do. A simple example is $\lfloor 1 - \{n\sqrt{2}\} \rfloor$, which is zero except when $n = 0$.

One does nonetheless feel that the discontinuities of τ_i are “mild”, as this function is continuous on that part of G/Γ which is identified with $(0, 1)^m$. However, the sequence $(g(n)\Gamma)_{n \in \mathbb{Z}}$ may well concentrate on a highly singular subset of G/Γ , as we discussed at length in [9]. Thus a certain amount of further work is required to obtain the expected result, which is the following.

Theorem 5.2 (Möbius and bracket polynomials). *Suppose that $p(n)$ is a bracket polynomial and that $\Psi : [0, 1] \rightarrow [-1, 1]$ is a Lipschitz function. Then we have the estimate*

$$\mathbb{E}_{n \in [N]} \mu(n) \Psi(\{p(n)\}) \ll_{A, \Psi} \log^{-A} N,$$

where the implied constant depends only on A , Ψ and the complexity of p (but is ineffective).

We shall illustrate how this theorem may be deduced from Theorem 1.1 by discussing two related special cases. We will then sketch the details that are required in order to write down a complete proof. The authors plan to include a complete proof of Theorem 5.2 in a future publication.

Both special cases will take place on the Heisenberg nilmanifold G/Γ , where

$$G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}, \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.$$

Computations with Mal'cev bases in this setting were given in [7, Appendix B] and then again in [9, §5], where we took

$$e_1 = \exp(X_1) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, e_2 = \exp(X_2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, e_3 = \exp(X_3) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We briefly recall some of the computations carried out in somewhat more detail in that paper; in any case the proofs are nothing more than computations with 3×3 matrices. The coordinate function $\psi : G \rightarrow \mathbb{R}^3$ is then given by the formula

$$\psi \left(\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \right) = (x, y, z - xy),$$

and the element written here is equivalent, under right multiplication by an element of Γ , to the element with coordinates

$$(\{x\}, \{y\}, \{z - xy - \lfloor x \rfloor y\}).$$

Note that this lies inside the fundamental domain $[0, 1)^3$. It follows that, for any $\alpha, \beta \in \mathbb{R}$, we have

$$\{n\beta \lfloor n\alpha \rfloor\} = \tau_3(g(n)\Gamma),$$

where $\tau_3 : G/\Gamma \rightarrow [0, 1)$ is the map into the third coordinate and $g : \mathbb{Z} \rightarrow G$ is the polynomial sequence given by

$$g(n) = \begin{pmatrix} 1 & n\alpha & n^2\alpha\beta \\ 0 & 1 & n\beta \\ 0 & 0 & 1 \end{pmatrix}.$$

This is an explicit example of the representation of a bracket polynomial, in this case $\{n\beta \lfloor n\alpha \rfloor\}$, in the form discussed in Bergelson and Leibman's theorem.

We discuss two different cases.

Case 1. $\alpha = \sqrt{2}$, $\beta = \sqrt{3}$. Then the sequence $(g(n)\Gamma)_{n \in [N]}$ is totally N^{-c} -equidistributed on G/Γ , which makes life rather easy. To prove the equidistribution one may use [9, Theorem 2.9] together with the lower bound

$$\min_{\substack{|k_1|, |k_2|, |k_3| \leq K \\ (k_1, k_2, k_3) \neq (0, 0, 0)}} \|k_1\sqrt{2} + k_2\sqrt{3}\|_{\mathbb{R}/\mathbb{Z}} \gg K^{-C},$$

which follows from the fact that, for any k_3 with $|k_3| \leq K$, $k_1\sqrt{2} + k_2\sqrt{3} + k_3$ satisfies a quartic over \mathbb{Z} with coefficients of size $K^{O(1)}$. Although the function τ_3 is not continuous, it is continuous outside of a subset of G/Γ of measure zero, namely outside of $[0, 1)^3 \setminus (0, 1)^3$. This means that it may be approximated by Lipschitz functions. More precisely, for any fixed Lipschitz function $\Psi : [0, 1] \rightarrow [-1, 1]$ and any $\varepsilon > 0$ one may find functions $F_1, F_2 : G/\Gamma \rightarrow \mathbb{C}$ with $\|F_1\|_\infty, \|F_2\|_\infty \leq 1$, $\|F_1\|_{\text{Lip}}, \|F_2\|_{\text{Lip}} \leq \varepsilon^{-O(1)}$, $|\Psi \circ \tau_3 - F_1| \leq F_2$ pointwise and $\int_{G/\Gamma} F_2 \leq \varepsilon$. From Proposition (2.1) we have

$$\mathbb{E}_{n \in [N]} \mu(n) F_1(g(n)\Gamma) \ll N^{-c},$$

and the uniform distribution of $(g(n)\Gamma)_{n \in [N]}$ implies that

$$\mathbb{E}_{n \in [N]} F_2(g(n)\Gamma) \leq \varepsilon + O(\varepsilon^{-O(1)} N^{-c}).$$

Now we have the bounds

$$\begin{aligned} |\mathbb{E}_{n \in [N]} \mu(n) \Psi(n\sqrt{3} \lfloor n\sqrt{2} \rfloor)| &= |\mathbb{E}_{n \in [N]} \mu(n) \Psi \circ \tau_3(g(n)\Gamma)| \\ &\leq |\mathbb{E}_{n \in [N]} \mu(n) F_1(g(n)\Gamma)| + \mathbb{E}_{n \in [N]} F_2(g(n)\Gamma). \end{aligned}$$

Letting $\varepsilon = N^{-c'}$ for some sufficiently small $c' > 0$, we obtain an effective and much stronger version of Theorem 5.2 in this case, namely the bound

$$\mathbb{E}_{n \in [N]} \mu(n) \Psi(\{n\sqrt{3} \lfloor n\sqrt{2} \rfloor\}) \ll N^{-c}.$$

Case 2. $\alpha = \beta = \sqrt{2}$. Now the sequence $(g(n)\Gamma)_{n \in [N]}$ is manifestly *not* uniformly distributed on G/Γ . In fact g takes values in the one-dimensional subgroup $G' \subseteq G$ defined by

$$G' = \left\{ \begin{pmatrix} 1 & x & x^2/2 \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}.$$

The preceding argument breaks down. One could appeal to Theorem 1.1 instead of Proposition 2.1, but the problem comes when one tries to control the term

$$\mathbb{E}_{n \in [N]} F_2(g(n)\Gamma).$$

Without knowing something more about the relation between the support properties of F_2 and the orbit $(g(n)\Gamma)_{n \in [N]}$, it is not possible to control this term.

In the case at hand $(g(n)\Gamma)_{n \in [N]}$ is N^{-c} -equidistributed in the nilmanifold G'/Γ' where $\Gamma' := \Gamma \cap G$. Topologically and algebraically this nilmanifold is nothing more than \mathbb{R}/\mathbb{Z} , but one should note carefully that the Haar measure on this nilmanifold is not the same as the measure induced from the Haar measure on G . This may be used to “explain” the observation that $n\sqrt{2}\lfloor n\sqrt{2} \rfloor$ is not uniformly distributed modulo one; see [2] for further details.

Inside G/Γ , G'/Γ' may be identified with the union of two segments

$$\left\{ \begin{pmatrix} 1 & x & x^{2/2} \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq x < 1 \right\} \cup \left\{ \begin{pmatrix} 1 & x & (1+x^2)/2 \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} : 0 \leq x < 1 \right\},$$

and this makes it clear that the induced map $\tau_3 : G'/\Gamma' \rightarrow [0, 1)$ is continuous away from a single point. By an analysis very similar to the preceding one it may once again be shown that

$$\mathbb{E}_{n \in [N]} \mu(n) \Psi(\{n\sqrt{2}\lfloor n\sqrt{2} \rfloor\}) \ll N^{-c}$$

for any fixed Lipschitz function $\Psi : [0, 1] \rightarrow [-1, 1]$.

Amongst examples of the form $n\beta\lfloor n\alpha \rfloor$ there is a third distinct case, typified by $\alpha = \beta = 2^{1/3}$. We leave the analysis of this to the reader.

Sketch proof of the general case of Theorem 5.2. By Theorem 5.1, the result of Bergelson and Leibman, it suffices to show, for any fixed Lipschitz function $\Psi : [0, 1] \rightarrow [-1, 1]$, that

$$\mathbb{E}_{n \in [N]} \mu(n) (\Psi \circ \tau_i)(g(n)\Gamma) \ll_A \log^{-A} N.$$

Here, the notation and parameters are as described in Theorem 5.1. Now τ_i is continuous outside the set $[0, 1)^m \setminus (0, 1)^m$, which has zero measure in G/Γ . The issue lies in understanding how the orbit $(g(n)\Gamma)_{n \in [N]}$ interacts with this.

Now the main results of [9] allow us to get a handle on this situation. Consider in particular the decomposition of g as $\varepsilon g' \gamma$ which was obtained in [9, Theorem 1.19]. Recall that $\varepsilon : \mathbb{Z} \rightarrow G$ is slowly varying, $\gamma : \mathbb{Z} \rightarrow G$ is rational and $g' : \mathbb{Z} \rightarrow G'$ is such that $(g'(n)\Gamma')_{n \in [N]}$ is totally equidistributed. For a full proof of Theorem 5.2 one would naturally need to specify appropriate quantitative parameters here. Suppose for simplicity that $\varepsilon = \gamma = \text{id}_G$ (this was, in fact, the case in the two examples above).

Choose a Mal'cev basis for G'/Γ' with coordinate map $\psi' : G' \rightarrow \mathbb{R}^{m'}$. Then G'/Γ' may be identified with the region $\psi'^{-1}([0, 1)^{m'}) \subseteq G$, and in this way we think of the coordinate function τ_i as a function on G'/Γ' . Write $\tilde{\tau}_i$ for the corresponding function on $[0, 1)^{m'}$. It can be shown, making extensive use of the results of [9, Appendix A], that $\tilde{\tau}_i$ is continuous outside of a *piecewise polynomial set* of positive codimension, that is to say outside of a finite union of sets each of which is defined by some polynomial inequalities $a \leq P(t_1, \dots, t_{m'}) < b$ and at least one nontrivial polynomial equation $Q(t_1, \dots, t_{m'}) = c$. Related matters are discussed at greater length in [2]; in the two examples we discussed, these piecewise polynomial sets were rather simple. These sets are certainly well-behaved enough that τ_i may be approximated using Lipschitz functions F_1 and F_2 as in our treatment of the bracket polynomial $n\sqrt{3}\lfloor n\sqrt{2} \rfloor$, and in this way one may use Theorem 1.1 to obtain the desired bound

$$\mathbb{E}_{n \in [N]} \mu(n) (\Psi \circ \tau_i)(g'(n)\Gamma) \ll_A \log^{-A} N.$$

If $G' \neq \{\text{id}\}$ then one may in fact use Proposition 2.1 to obtain the stronger bound of N^{-c} , as in the examples.

If ε and γ are not trivial it is even more complicated to write down a fully rigorous argument, but conceptually things are not much harder at all. The introduction of the smooth function $\varepsilon(n)$ has a rather benign effect; if n ranges over an interval of length $\delta'N$, for suitably small $\delta' = \delta'(\delta)$, the discontinuities of the functions $x \mapsto \tau_i(\varepsilon(n)x\Gamma)$ are all contained inside a “nice” set of measure at most δ , and one may proceed much as before. All one need do, then, is split the range $[N]$ into suitably short intervals of this type.

The introduction of γ may be handled much as it was in the proof of Theorem 1.1. One splits each of the intervals from the previous paragraph into progressions P_j with the same (small) common difference q such that $\gamma(n)\Gamma$ is constant and equal to $\gamma_j\Gamma$ on P . One then works with the conjugated sequences $\gamma_j^{-1}g'(n)\gamma_j$ as we did at the end of §2. \square

We conclude by remarking on some variants and generalizations of Theorem 5.2. If p_1, \dots, p_M are bracket polynomials and $F : (\mathbb{R}/\mathbb{Z})^M \rightarrow \mathbb{C}$ is a smooth function then one could establish the estimate

$$\mathbb{E}_{n \in [N]} \mu(n) F(\{p_1(n)\}, \dots, \{p_M(n)\}) \ll_A \log^{-A} N$$

by Fourier decomposition of F and Theorem 5.2. One could, if desired, restrict the range of the average to some fixed subprogression $P \subseteq [N]$ by the standard technique of approximating the cutoff $1_P(n)$ by a smoother function $\tilde{1}_P(n)$ and then developing this as a Fourier expansion.

6. THE LIOUVILLE FUNCTION

Everything we have proved for the Möbius function also holds for the Liouville function $\lambda : \mathbb{N} \rightarrow \{-1, 1\}$, defined to be the unique completely multiplicative function such that $\lambda(p) = -1$ for all primes p . This function is related to the Möbius function via the identity

$$\lambda(n) = \sum_{r:r^2|n} \mu(n/r^2).$$

Thus, with the notation and assumptions of Theorem 1.1, we have

$$|\mathbb{E}_{n \in [N]} \lambda(n) F(g(n)\Gamma)| \ll \sum_{1 \leq r \leq \sqrt{N}} \frac{1}{r^2} |\mathbb{E}_{m \in [N/r^2]} \mu(m) F(g(r^2 m)\Gamma)|.$$

Now by [9, Corollary 6.8] $m \mapsto g(r^2 m)$ is a polynomial sequence with coefficients in the same filtration G_\bullet as g , and so we have the bound

$$|\mathbb{E}_{m \in [N/r^2]} \mu(m) F(g(r^2 m)\Gamma)| \ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \log^{-A}(N/r^2)$$

uniformly in r , so long as $N/r^2 \geq 2$. Summing over r we obtain

$$\begin{aligned} |\mathbb{E}_{n \in [N]} \lambda(n) F(g(n)\Gamma)| &\ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \left(\sum_{r \leq \sqrt{N}/2} \frac{1}{r^2} \log^{-A}(N/r^2) \right. \\ &\quad \left. + \sum_{\sqrt{N}/2 < r \leq \sqrt{N}} \frac{1}{r^2} \right) \\ &\ll_{m,d,A} Q^{O_{m,d,A}(1)} (1 + \|F\|_{\text{Lip}}) \log^{-A} N. \end{aligned}$$

This is precisely Theorem 1.1, but with λ taking the place of μ . In a similar fashion, all of the results of the preceding section concerning bracket polynomials may now also be deduced with λ in place of μ .

7. A RECURRENCE RESULT ALONG THE PRIMES

In this section we derive the following result. Here p_1, p_2, p_3, \dots is the sequence of primes.

Theorem 7.1 (Prime return times on a nilmanifold). *Suppose that G/Γ is a nilmanifold and that $g \in G$ is such that left-multiplication by g is ergodic. Then for every $x \in G/\Gamma$ the sequence $(g^{p_n} x \Gamma)_{n=1,2,\dots}$ is equidistributed in G/Γ in the sense that*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} F(g^{p_n} x \Gamma) = \int_{G/\Gamma} F$$

for all continuous functions $F : G/\Gamma \rightarrow [-1, 1]$.

Remarks. We recall (from discussions in the companion paper [9]) Leon Green's criterion for ergodicity of left-multiplication by g ; this map is ergodic if and only if rotation by $\pi(g)$ is ergodic on the horizontal torus $(G/\Gamma)_{\text{ab}}$, that is to say if and only if the entries of $\pi(g)$ together with 1 are linearly independent over \mathbb{Q} . If this is the case then left-multiplication by any power of g is uniquely ergodic, that is to say

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} F(g^{t_n} x \Gamma) = \int_{G/\Gamma} F \quad (7.1)$$

for all $x \in G/\Gamma$ and for $t = 1, 2, 3, \dots$

Proof of Theorem 7.1. Let w be a large number and set $W := \prod_{p \leq w} p$. Fix a nilmanifold G/Γ and a continuous (and hence Lipschitz) function $F : G/\Gamma \rightarrow [-1, 1]$. Then uniformly in the residues b coprime to W we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \left(\frac{\phi(W)}{W} \Lambda'(Wn + b) - 1 \right) F(g^n x \Gamma) = o_{w \rightarrow \infty}(1), \quad (7.2)$$

where the convergence is uniform in $x \in G/\Gamma$ and $g \in G$. This follows very quickly from [8, Proposition 10.2], which was proved under the assumption of the Möbius and Nilsequences conjectures $\text{MN}(s)$ which we have established in this paper. Recall that $\Lambda'(p) = \log p$ and that $\Lambda'(n) = 0$ if n is not a prime, that is to say Λ' is a modified version of the von Mangoldt function with no support on the prime powers p^2, p^3, \dots . We recall that the proof of (7.2) is quite substantial. One splits the von Mangoldt function Λ in a certain way as the sum of two pieces $\Lambda^\sharp + \Lambda^\flat$. The contribution from the second piece is bounded using the $\text{MN}(s)$ conjecture, and this is not particularly difficult. The

contribution from the first piece is bounded using the machinery of Gowers norms, and here one must estimate the dual Gowers norm of the nilsequence $F(g^n x \Gamma)$ as well as the Gowers norm of objects related to Λ^\sharp . This is a substantial amount of work.

Let us return to the proof at hand. Since (7.2) is uniform in g and x , we may replace g by g^W and x by $g^b x$ to get

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} \left(\frac{\phi(W)}{W} \Lambda'(n) - 1 \right) F(g^n x \Gamma) = o_{w \rightarrow \infty}(1)$$

uniformly for all progressions $P_{b,W} = \{Wn + b : n \in [N]\}$, $b = 0, 1, \dots, W-1$. However it follows from (7.1) that, for fixed b and W ,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} F(g^n x \Gamma) = \int_{G/\Gamma} F.$$

Comparing these last two expressions we obtain

$$\frac{\phi(W)}{W} \lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F + o_{w \rightarrow \infty}(1),$$

uniformly for b coprime to W . Now if b is not coprime to W we obviously have

$$\frac{\phi(W)}{W} \lim_{N \rightarrow \infty} \mathbb{E}_{n \in P_{b,W}} \Lambda'(n) F(g^n x \Gamma) = o_{w \rightarrow \infty}(1)$$

since Λ' is supported on the primes and F is bounded by 1.

Summing over b , one may conclude that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [WN]} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F + o_{w \rightarrow \infty}(1).$$

This is easily seen to imply that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F + o_{w \rightarrow \infty}(1).$$

The left-hand side no longer depends on w , so we may let $w \rightarrow \infty$. Doing so, we obtain

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \Lambda'(n) F(g^n x \Gamma) = \int_{G/\Gamma} F.$$

An easy argument using the prime number theorem, noting that $\Lambda'(p_n)$ is essentially $\log N$ for almost all primes p_n , $n \leq N$, concludes the proof. \square

Very straightforward approximation arguments allow one to replace the continuous function F by a function with mild discontinuities. In this way one could prove, for example, that the sequence $p_n \sqrt{3} \lfloor p_n \sqrt{2} \rfloor$ is uniformly distributed modulo one. We leave the details, which are essentially all present in the earlier discussion of $n \sqrt{3} \lfloor n \sqrt{2} \rfloor$, to the reader.

APPENDIX A. MÖBIUS AND PERIODIC FUNCTIONS

In this appendix we give the proof of Proposition A.2. The argument is, quite apart from being completely standard, already contained in [7, Chapter 3]. We nonetheless take the opportunity to recall it here, as we wish to emphasise the fact that the main

input to this part of the argument is information on the zeros of L -functions. Our starting point is the following proposition.

Proposition A.1. *For any $A > 0$ we have*

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{\chi(n)} \ll_A q^{1/2} \log^{-A} N \quad (\text{A.1})$$

for all Dirichlet characters χ to modulus q .

Remark. This follows from the nonexistence of zeros of $L(s, \chi)$ close to the line $\Re s = 1$. For the details, see [10, Prop. 5.29]. As noted in [10, p. 124] there are difficulties involved in applying the standard Perron's formula approach to $\mathbb{E}_{n \in [N]} \mu(n) \chi(n)$ directly, and it is rather easier to first obtain bounds on $\mathbb{E}_{n \in [N]} \Lambda(n) \chi(n)$.

Using standard techniques of harmonic analysis we may obtain the following consequence of Proposition A.1.

Proposition A.2 (Möbius is orthogonal to periodic sequences). *Let $f : \mathbb{N} \rightarrow \mathbb{C}$ be a sequence bounded in magnitude by 1 which is periodic of some period $q \geq 1$. Then we have*

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} \ll_A q \log^{-A} N$$

for all $A > 0$, where the implied constant is ineffective.

Proof. We first establish the estimate under the additional assumption that $f(n)$ vanishes whenever $(n, q) \neq 1$. Then f can be viewed as a function on the multiplicative group $(\mathbb{Z}/q\mathbb{Z})^\times$, and thus has a Fourier expansion

$$f(n) = \sum_{\chi} \hat{f}(\chi) \chi(n), \text{ where } \hat{f}(\chi) := \mathbb{E}_{n \in (\mathbb{Z}/q\mathbb{Z})^\times} f(n) \overline{\chi(n)},$$

with χ ranging over all the characters on $(\mathbb{Z}/q\mathbb{Z})^\times$. Applying Proposition A.1 and the triangle inequality, we conclude

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} \ll_A q^{1/2} \log^{-A} N \left(\sum_{\chi} |\hat{f}(\chi)| \right).$$

But from Cauchy-Schwarz and Plancherel we have

$$\sum_{\chi} |\hat{f}(\chi)| \leq \phi(q)^{1/2} \left(\sum_{\chi} |\hat{f}(\chi)|^2 \right)^{1/2} = \phi(q)^{1/2} (\mathbb{E}_{n \in (\mathbb{Z}/q\mathbb{Z})^\times} |f(n)|^2)^{1/2} = O(\phi(q)^{1/2}),$$

where $\phi(q) := |(\mathbb{Z}/q\mathbb{Z})^\times|$ is the Euler totient function. Since $\phi(q) \leq q$, the claim follows.

Now we consider the general case, in which (n, q) is not necessarily equal to 1 on the support of f . Observe that if $\mu(n)$ is non-zero, then n is square-free, and we can split $n = dm$, where $d = (n, q)$ is square-free (so $\mu^2(d) = 1$) and m is coprime to q . Furthermore we have $\mu(n) = \mu(d)\mu(m)$. We thus obtain the decomposition

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} = \frac{1}{N} \sum_{d|q; \mu^2(d)=1} \mu(d) \sum_{1 \leq m \leq N/d} \mu(m) \overline{f(dm)} 1_{(m,q)=1}. \quad (\text{A.2})$$

The sequence $m \mapsto f(dm) 1_{(m,q)=1}$ is periodic of period q/d and vanishes whenever $(m, q/d) \neq 1$, hence by the preceding arguments

$$\sum_{1 \leq m \leq N/d} \mu(m) \overline{f(dm)} 1_{(m,q)=1} \ll_A \frac{Nq}{d^2} \log^{-A} N.$$

Thus from (A.2) we have

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{f(n)} \ll_A q \log^{-A} N \sum_{d|q} \frac{1}{d^2} \ll q \log^{-A} N,$$

concluding the proof of Proposition A.2. \square

APPENDIX B. ON TOTAL EQUIDISTRIBUTION OF NILSEQUENCES

During the proof that Proposition 2.1 implies Theorem 1.1 (specifically, in the claim towards the end of §2), we implicitly used the following technical fact⁴.

Lemma B.1. *Suppose that G/Γ is an m -dimensional nilmanifold, and suppose that G_\bullet is a filtration of degree d . Suppose that G/Γ has a Q -rational Mal'cev basis \mathcal{X} adapted to the filtration G_\bullet . Let G' be a rational subgroup of G which comes together with a Mal'cev basis \mathcal{X}' for G'/Γ' (where $\Gamma' = G' \cap \Gamma$), in which each element is a Q -rational combination of the elements of \mathcal{X} .*

Let $g' \in \text{poly}(Z, G_\bullet)$ be a polynomial sequence taking values in G' and suppose that $(g'(n)\Gamma')_{n \in [N]}$ is totally δ -equidistributed in G'/Γ' .

Then for every Q -rational $\gamma \in G$ with coordinates bounded by Q the following is true. The conjugate $G'_\gamma := \gamma^{-1}G'\gamma$ is a rational subgroup of G which comes together with a Mal'cev basis \mathcal{X}'_γ for G'_γ/Γ'_γ (where $\Gamma'_\gamma := G'_\gamma \cap \Gamma$), in which element is a $Q^{O_{m,d}(1)}$ -rational combination of the elements of \mathcal{X} . Furthermore, the conjugate sequence $g'_\gamma := \gamma^{-1}g'\gamma$, which manifestly takes values in G'_γ , is $Q^{C_{m,d}}\delta^{C_{m,d}}$ -totally equidistributed in G'_γ/Γ'_γ .

Proof. This would follow quite straightforwardly from the results in [9, Appendix A] if, instead of Γ'_γ , we worked with $\gamma^{-1}\Gamma'\gamma$. The result as stated follows from this, the lemma below and the fact that $\gamma^{-1}\Gamma'\gamma$ and Γ'_γ are $Q^{O_{m,d}(1)}$ -commensurable. This follows from the fact that $[\Gamma : \Gamma \cap \gamma^{-1}\Gamma'\gamma] \leq Q^{O_{m,d}(1)}$, which may in turn be proven by noting that the \mathcal{X} coordinates of every element in $\gamma^{-1}\Gamma'\gamma$ are rationals over some fixed denominator $Q^{O_{m,d}(1)}$, as follows from [9, Appendix A]. \square

Lemma B.2. *Suppose that G is an m -dimensional simply-connected nilpotent Lie group. Let Γ_1, Γ_2 be two uniform subgroups which are $\frac{1}{\delta}$ -commensurable in the sense that $\Gamma_1 \cap \Gamma_2$ has index at most $\frac{1}{\delta}$ in both Γ_1 and Γ_2 . Suppose that there is a $\frac{1}{\delta}$ -rational Mal'cev basis \mathcal{X}_1 for G/Γ_1 , and that $g : \mathbb{Z} \rightarrow G$ is a polynomial sequence of degree d with the property that $(g(n)\Gamma_1)_{n \in [N]}$ is totally δ -equidistributed with respect to the metric $d_{\mathcal{X}_1}$. Then there is a Mal'cev basis \mathcal{X}_2 for G/Γ_2 , each element of which is a $\delta^{-O_{m,d}(1)}$ -rational combination of the elements of \mathcal{X}_1 , and such that $(g(n)\Gamma_2)_{n \in [N]}$ is totally $\delta^{C_{m,d}}$ -equidistributed with respect to the metric $d_{\mathcal{X}_2}$.*

Remark. This lemma would be false without the assumption of *total* equidistribution. For example, the sequence $(n/N)_{n \in [N]}$ is highly equidistributed on \mathbb{R}/\mathbb{Z} , but not at all close to equidistributed on $\mathbb{R}/2\mathbb{Z}$, even though the lattices $\mathbb{Z}, 2\mathbb{Z}$ are highly commensurable.

⁴We note that this fact was absent in our original posting of this paper on the ArXiv in 2008; we apologize for this oversight.

Proof. By applying the lemma twice (first with Γ_1 and $\Gamma_1 \cap \Gamma_2$ and then with $\Gamma_1 \cap \Gamma_2$ and Γ_2), we may reduce to the case where one of Γ_1, Γ_2 is contained in the other. The existence of \mathcal{X}_2 follows, once again, from the material in [9, Appendix A]. The commensurability of Γ_1 and Γ_2 easily implies that the coordinates of Γ_2 with respect to \mathcal{X}_1 are all $\delta^{-O_{m,d}(1)}$ -rational.

It remains to establish the statement about total equidistribution. If $\Gamma_1 \subseteq \Gamma_2$ then the fundamental domain for G/Γ_1 is just a union of $\frac{1}{\delta}$ copies of that for G/Γ_2 , and so the result is clear. We may assume, then, that $\Gamma_2 \subseteq \Gamma_1$. Applying the quantitative Leibman dichotomy, namely [9, Theorem 1.16], this may be further reduced to the corresponding abelian statement. Indeed that result immediately implies that $(g(n)\Gamma)_{n \in [N]}$ is totally δ -equidistributed if and only if $(\pi(g(n)\Gamma)_{n \in [N]})$ is totally δ' -equidistributed in the horizontal torus $(G/\Gamma)_{\text{ab}}$, where the dependence between δ and δ' in both directions of this implication is polynomial of degree $O_{m,d}(1)$.

In the abelian case we have an isomorphism between $G/[G, G]$ and $\mathbb{R}^{m_{\text{ab}}}$, and we may change coordinates so that $G/[G, G]\Gamma_1 = \mathbb{Z}^{m_{\text{ab}}}$. In this setting $G/[G, G]\Gamma_2$ is then the image of $\mathbb{Z}^{m_{\text{ab}}}$ under some invertible linear transformation $A \in \text{GL}_{m_{\text{ab}}}(\mathbb{Z})$ with matrix coefficients of size $\delta^{-O_{m,d}(1)}$. We have reduced matters, then, to the following assertion.

Lemma B.3. *Let $m \geq 1$ be a positive integer and suppose that $A \in \text{GL}_m(\mathbb{Z})$ is invertible and has entries bounded by $\frac{1}{\delta}$. Let $g : \mathbb{Z} \rightarrow \mathbb{R}^m$ be a polynomial of degree d , and write $g' := A \circ g$. Suppose that the sequence $(g'(n)\mathbb{Z}^m)_{n \in [N]}$ is totally δ -equidistributed in $\mathbb{R}^m/\mathbb{Z}^m$. Then the sequence $(g(n)\mathbb{Z}^m)_{n \in [N]}$ is totally δ' -equidistributed in $\mathbb{R}^m/\mathbb{Z}^m$, for some $\delta' = \delta^{c_{m,d}}$.*

Proof. We use the quantitative Leibman dichotomy, [9, Theorem 1.16], but now of course in the abelian setting. If $(g(n)\mathbb{Z}^m)_{n \in [N]}$ fails to be totally δ' -equidistributed then there is some $k \in \mathbb{Z}^m$, $|k| \leq \delta'^{-O_{m,d}(1)}$, such that $\|k \cdot g\|_{C^\infty[N]} \leq \delta^{-O_{m,d}(1)}$. This implies that $\|k' \cdot g'\|_{C^\infty[N]} \leq \delta'^{-O_{m,d}(1)}$, where $k' := (\det A)kA^{-1}$; note that $k' \in \mathbb{Z}^m$ and that $|k'| \leq \delta'^{-O_{m,d}(1)}$. Recalling the definition of $\|\cdot\|_{C^\infty[N]}$ and passing to an appropriate subprogression $P \subseteq [N]$, one sees that $(g(n))_{n \in P}$ concentrates near the set $\{x \in \mathbb{R}^m/\mathbb{Z}^m : \|k \cdot x\|_{\mathbb{R}/\mathbb{Z}} < \frac{1}{10}\}$. This contradicts the assumed total δ -equidistribution of $(g'(n)\mathbb{Z}^m)_{n \in [N]}$, provided that δ' is chosen to be a sufficiently small power of δ . \square

REFERENCES

- [1] V. Bergelson and I. J. Håland, *Sets of recurrence and generalized polynomials*, Convergence in ergodic theory and probability (Columbus, OH, 1993), 91–110, Ohio State Univ. Math. Res. Inst. Publ., **5**, de Gruyter, Berlin, 1996.
- [2] V. Bergelson and A. Leibman, *Distribution of values of bounded generalized polynomials*, Acta Mathematica **198** (2007), 155–230.
- [3] H. Davenport, *On some infinite series involving arithmetical functions. II*, Quart. J. Math. Oxf. **8** (1937), 313–320.
- [4] B. J. Green, *Generalising the Hardy-Littlewood method for primes*, International Congress of Mathematicians. Vol. II, 373–399, Eur. Math. Soc., Zürich, 2006.
- [5] ———, *Three topics in additive prime number theory*, Current developments in mathematics, 2007, 1–41, Int. Press, Somerville, MA, 2009.
- [6] B. J. Green and T. C. Tao, *An inverse theorem for the Gowers U^3 -norm*, Proc. Edinburgh Math. Soc. **51** (2008), no. 1, 73–153.

- [7] ———, *Quadratic uniformity of the Möbius function*, Ann. Inst. Fourier (Grenoble) **58** (2008), no. 6, 1863–1935.
- [8] ———, *Linear equations in primes*, to appear in Annals of Math.
- [9] ———, *The quantitative behaviour of polynomial orbits on nilmanifolds*, to appear in Annals of Math.
- [10] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, **53**. American Mathematical Society, Providence, RI, 2004. xii+615 pp
- [11] A. Leibman, *Polynomial sequences in groups*, Journal of Algebra **201** (1998), 189–206.
- [12] ———, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory and Dynamical Systems **25** (2005), no. 1, 201–213.
- [13] A. Mal'cev, *On a class of homogeneous spaces*, Izvestiya Akad. Nauk SSSR, Ser Mat. **13** (1949), 9–32.
- [14] T. C. Tao, *Obstructions to uniformity, and arithmetic patterns in the primes*, Pure Appl. Math. Q. **2** (2006), no. 2, part 2, 395–433.
- [15] ———, *The dichotomy between structure and randomness, arithmetic progressions, and the primes*, International Congress of Mathematicians. Vol. I, 581–608, Eur. Math. Soc., Zürich, 2007.
- [16] R. C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A-B **285** (1977), no. 16, A981–A983.
- [17] ———, *The Hardy-Littlewood method*, Second edition. Cambridge Tracts in Mathematics, **125**. Cambridge University Press, Cambridge, 1997. xiv+232 pp

CENTRE FOR MATHEMATICAL SCIENCES, WILBERFORCE ROAD, CAMBRIDGE CB3 0WA, ENGLAND

E-mail address: `b.j.green@dpmms.cam.ac.uk`

UCLA DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA 90095-1596, USA

E-mail address: `tao@math.ucla.edu`